The Dissertation Committee for Stephanie Nicole Baker certifies that this is the approved version of the following dissertation:

Post Hoc Discernment of Developmental Mathematics Noncognitive Factors and Concept Transfer

Committee:

Philip Uri Treisman, Supervisor

Catherine Riegle-Crumb

Michael Starbird

Victor Saenz

Post Hoc Discernment of Developmental Mathematics Noncognitive Factors and Concept Transfer

by

Stephanie Nicole Baker, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin
December 2016

Dedication

I dedicate this dissertation to my friends and family for their love and for always believing in me. I am lucky to know each and every one of you.

Acknowledgements

This has been an incredibly powerful, yet difficult, journey that was supported by numerous individuals. If I were to thank all of the amazing people who helped me along the way, this section would be longer than the entire dissertation. Please know I appreciate even the littlest ways in which many have assisted.

I must first thank the participating institutions, their helpful faculty and staff, and the study participants. You are the reason I was able to complete this study and I am grateful.

To my committee: Most importantly, thank you for your acceptance and friendship. I believe you all truly care about your craft and care about the impact it has on others. I thank you immensely for your continued patience. Even when I thought about giving up, you never gave up on me, and this motivated me to carry on.

Dr. Uri Treisman, I will never be able to adequately describe how much your advice and words of encouragement have meant to me. We explored so many dissertation topics and I am grateful you guided me to a topic about which I am incredibly passionate. You encouraged me to persist and were attentive to my wellbeing as more than a researcher, but as a person. As a bonus, I now have memorable quotes from our meetings, such as: “Nothing says doctor like a kazoo.” and “Ten is the new nine.”

Dr. Catherine Riegle-Crumb, thank you for always making time to meet with me and allowing me to discuss my frustrations and fears. Thank you for pushing me to critically examine my work, especially when I asked for feedback (which you graciously provided) as I modified the place value assessment.

Dr. Michael Starbird, thank you for asking thought-provoking questions and requiring me to justify my positions and think outside of the box. Thank you, also, for helping me keep the larger picture in mind as I focused on smaller pieces in my study.

Dr. Victor Saenz, thank you for your willingness to be on my committee without substantial information about my background. It was immediately apparent to me how lucky I was to have you join and I appreciate all of the feedback you provided as I began my study.

Dr. Taylor Martin, thank you for inviting me to join the Active Learning Lab and immediately immersing me in the world of STEM education research, especially as it relates to quantitative methods. Even after departing The University of Texas at Austin, I knew I could rely on you to help guide me as I encountered difficult situations. Thank you for allowing me to work on diverse projects and for expecting publications. While those late nights were sometimes overwhelming, the end results always made the experiences worthwhile.

Dr. Leema Berland, thank you for helping me see, contrary to the popular saying, sometimes the simplest solution is not always the correct one. You pushed me to improve my skills as a qualitative researcher and helped me formulate meaningful questions.

To all of the STEM Education faculty at The University of Texas at Austin, thank you for believing in me and allowing me to continue to pursue this work despite my taking longer than expected. Specifically, thank you Dr. Walter Stroup for inspiring me in the first education course I took at UT. You completely changed my thinking about learning and the ways to assess it. You are the reason I entered the Math Education Doctoral Program—yes, it's your fault—and I am grateful. Dr. Jill Marshall, thank you for regularly checking with me on my progress and offering advice.

I am forever grateful to the late Dr. Miguel Paredes, my advisor who oversaw my mathematics research as I pursued my master's degree. He was a good friend who treated me like a colleague and encouraged me to pursue my doctoral degree. I will always miss the times we spent talking about group theory, sometimes four or five hours at a time.

Thank you to my friends and family who have supported me through this long process. I kept promising I would eventually finish and rejoin the world and now the time has finally arrived.

Dad and Mom Baker, you helped me in so many ways throughout this process; thank you for continuously supporting my goals and assisting me in achieving them. Dad and Mom Botto, thank you also for your neverending support; your advice and words of inspiration helped tremendously. Mom, thank you for trying to help me find that one perfect word; it was reminiscent of our dictionary days. Thank you to my sisters, brother, and Momma2 for remaining steadfast in my corner.

Thank you, Rory, for your loving support and encouragement. I could not possibly have done this without you. You have been my partner through thick and thin. You let me cry on your shoulder when I thought it was too tough to continue and you reinvigorated me. You listened when I explained minute details of my study that are probably only interesting to me, and you helped me decipher confusing aspects of my work.

Eugenie, your friendship has meant the world to me. I can hardly believe we have been friends for twelve years! You are an excellent life partner! Thank you to the Peacock family for your continued support. Linda P., as you suggested, I finally "wrote the thing." Thank you to all of you who helped me by taking various iterations of the place value assessment, especially those of you who took the longest version. Your aid substantially contributed to the final product.

I am appreciative of an overwhelming number of members of the UT community who I also consider friends. I sincerely thank former and current Charles A. Dana Center staff, especially Lilly S., Rachele S., Francesca F. L., Rachel J., Jennifer D., and Rahel K., for welcoming me into the fold and assisting whenever possible. Jennifer D., I particularly thank you for your assistance with data collection, troubleshooting, and enjoyable conversations. I am grateful to the statistical consultants in UT's Department of Statistics and Data Sciences. Erika H., I can never thank you enough! Your statistical expertise helped me immensely and, because I wanted to make certain to address every nook and cranny, I likely visited you more than any past client. You have become a true friend.

Thank you to all of the members of the Active Learning Lab. Vanessa S., Bill M., Chris G., Tom B., and Stephanie R., you are wonderful friends to have as colleagues; thank you for pushing my thinking forward. Carmen P.S., our laughter was nonstop and exhilarating; thank for your friendship and smiles. Pat K., my comrade, partner in crime, co-author, and office bud, thank you for your friendship and helping me tackle both methodological and logistical research problems, especially when food or coffee was involved. Sarah H. B., I admire you more than you will ever know. I was proud to be your sous chef, and I learned so much from you. I appreciate all the time we spent discussing research and how it led to us becoming good friends. Joey H., your assistance, especially in the final hours, was of tremendous help. I can never fully repay you, but I will certainly try. Now, it's your turn, so make sure you know where your towel is at all times.

Thank you to all of my friends in the math department and the extended math department family. Nick R., thank you for responding to my initial email about a study group and for continuing to be a great friend. Mark N., thank you also for staying a true friend even after moving to a place without burger night. Cyntreva P., thank you for helping with early data collection and organization; I thoroughly enjoyed working with you.

Lastly, I am indebted to the researchers whose contributions in their field most notably impacted my work. My ideas about meaningful and efficient ways to measure noncognitive factors were appreciably influenced by the practical measurement work of Dr. David Yeager and his colleagues, as well as the retrospective pretest research by Dr. George Howard. I thank Dr. Giyoo Hatano, Dr. Harry Broudy, Dr. Daniel Schwartz, Dr. John Bransford, and Dr. David Sears for their inspiring perspectives on concept transfer. I thank Dr. Tracy Rusch and Dr. Mary Hannigan for creating the Assessment of Place Value Understanding and providing a framework with which to assess explicit understanding.

Post Hoc Discernment of Developmental Mathematics Noncognitive Factors and Concept Transfer

Stephanie Nicole Baker, Ph.D.

The University of Texas at Austin, 2016

Supervisor: Philip Uri Treisman

One purpose of this study was to determine if students in a non-traditional developmental mathematics course improved on five developmental mathematics noncognitive factors—math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness—believed to be relevant to student success. I also examined if changes in these factors predicted course achievement. Another purpose was to explore whether or not Foundations students would transfer their knowledge to place value problems involving varied bases and contexts. A final purpose was to investigate the utility of then-surveys that retrospectively measure participants' pre-intervention noncognitive factors.

In response to policy pressures to increase completion rates, community colleges are experimenting with research-based strategies that create demand for learning, increase students' competence valuation, and improve their productive persistence. The New Mathways Project's Foundations of Mathematical Reasoning course is built around one such strategy.

In this exploratory study (N = 597), I investigated the impact of using Foundations on the development of students' noncognitive factors and on mathematical success. My student measures included: pre-post-then-surveys of noncognitive factors,

math course grades, math final exam grades, percent attendance, a place value assessment of transfer, and one-on-one interviews. I used multilevel models to analyze my quantitative research questions and created evidence markers for qualitative analysis of the transfer assessment. I conducted interviews to provide additional insight.

Students significantly improved their math equanimity, but had stable, mid-range scores on the other factors. Positive changes in math self-efficacy and low initial math equanimity were associated with higher grades. Pre-surveys of equanimity may be more accurate than then-surveys, but pre-surveys of math mindset, math self-efficacy, and math belongingness may be interchangeable with then-surveys. Contrary to popular findings, the then-surveys did not provide larger estimates of program effects than pre-surveys. Overall, students evidenced minimal transfer. Interviewees exhibited greater changes in noncognitive factors and evidenced more transfer than other students.

This study provides valuable information for the potential users of the NMP materials. It contributes to, and points out complications with, transfer research. Lastly, it adds to research on retrospective measures, which are rarely used in mathematics education research.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

RATIONALE

This is a period of tremendous change in community college education. The 21st-Century Commission on the Future of Community Colleges was tasked with taking a critical look at the difficulties and opportunities facing community colleges and they received feedback from 1,300 stakeholders, including students, college faculty and staff, policy makers, and college presidents (American Association of Community Colleges, 2012). In broad overview, the Commission argued that if the open door mission of community colleges is to persist, then everything else about these access-focused institutions must change. Community colleges, the Commission asserted, have been remarkably effective at democratizing access to higher education for low-income students and students of color. And, they have accomplished this goal at relatively low cost. With no additional revenue, these institutions are now being called upon to become engines of degree completion.

Until recently, state and local funding of community colleges has been based almost exclusively on enrollment. Between 2008 and 2015 the majority of states have redesigned their funding formulas to reward the completion of certificates, licenses, and degrees with labor market value. This policy shift is at the heart of a national movement to increase the number and proportion of adults obtaining higher education certifications.

The motivation for this near ubiquitous policy change is economic and grounded in concerns about economic competitiveness. In the next three years, approximately two thirds of jobs in America will require job-aligned education beyond secondary school, but the United States has fallen as a leader in degree completion to 16th in the world for 25-34-year-olds (American Association of Community Colleges, 2012). In 2012, the

American Association of Community Colleges (AACC) recommended adding 20 million workers to the workforce who have completed a postsecondary education by 2027 to drastically lessen economic, racial, ethnic, and gender inequalities. The central issue is not that students aren't entering postsecondary programs; it is that students are not *completing* postsecondary programs.

The odds are stacked against mainstream community college students, as approximately 70% of the incoming students are required to take mathematics, English, and/or reading developmental courses[1] prior to being considered college ready and only one fourth of those students receive a terminal degree, certificate or license within eight years (Achieving the Dream et al., 2015b).[2] Attention to degree completion coupled with the fact that the majority of developmental education students are enrolled in developmental mathematics courses has put mathematics in the spotlight. Since community college students must generally complete at least one credit-bearing mathematics course, developmental mathematics coursework can delay or even prevent them from earning a post-secondary degree. In 2012, Texas' non-developmental education students were 50% more likely than their developmental education counterparts to complete a degree that is useful in the workforce or transfer into a baccalaureate program and, nationally, developmental math students had a 59% failure rate (The Texas Association of Community Colleges, 2012). As a result, developmental mathematics is the "primary barrier for students ever being able to complete a post-secondary degree" (Stigler, Givvin, & Thompson, 2010, p. 2).

---

[1] In the past, developmental courses were referred to as remedial courses.

[2] Complete College America estimates that only one tenth of students in developmental courses graduate within six years (Vandal, 2015).

The aforementioned challenges facing community colleges are obstructing paths to success and this creates an especially inequitable situation for students who could benefit the most from community colleges—minorities and students who are socioeconomically disadvantaged. The majority of community college students are non-traditional (e.g., older, manage a household), members of ethnic or racial minorities, economically disadvantaged, and/or first generation college students (Yeager, Bryk, Muhich, Hausman, & Morales, 2013).

Johnstone (2015) provides an example of how income gaps translate to education gaps: When students who scored in the middle range on the SAT were from the highest income quartile, they were four times more likely to obtain a degree by age 24 than students from the lowest income quartile with scores in the same range. With top-level SAT performers, "[t]he highest level income quartile achieves a college degree 82% of the time by age 24, while those in the lowest income quartile do so just 44% of the time" (p. 7). Similar trends are found with minority students. This data shines light on the false assumption that ability is the source of the gap between students who do and students who do not have ultimate postsecondary success.

The AACC (American Association of Community Colleges, 2012, p. vii) asserts that "the American dream [of intergenerational upward mobility] is imperiled" and "community colleges can help reclaim that dream", but "stepping up to this challenge will require dramatic redesign of these institutions, their mission, and most critically, their students' educational experiences."

In light of all of these issues, institutions are rethinking remediation in the context of a broader student success strategy. Six prominent organizations[3] that are dedicated to

[3] The joint statement was produced by the following organizations: Achieving the Dream, American Association of Community Colleges, Charles A. Dana Center of The University of Texas at Austin, Complete College America, Education Commission of the States, and Jobs for the Future.

research and reform in developmental education recently released a joint statement about this strategy, which involves six core principles:

1. Every student's postsecondary education begins with an intake process to choose an academic direction and identify the support needed to pass relevant credit-bearing gateway courses in the first year.
2. Enrollment in college-level math and English courses or course sequences aligned with the student's program of study is the default placement for the vast majority of students.
3. Academic and nonacademic support is provided in conjunction with gateway courses in the student's academic or career area of interest though co-requisite or other models with evidence of success in which supports are embedded in curricula and instructional strategies.
4. Students for whom the default college-level course placement is not appropriate, even with additional mandatory support, are enrolled in rigorous, streamlined remediation options that align with the knowledge and skills required for success in gateway courses in their academic or career area of interest.
5. Every student is engaged with content of required gateway courses that is aligned with his or her academic program of study—especially in math.
6. Every student is supported to stay on track to a college credential, from intake forward, through the institution's use of effective mechanisms to generate, share, and act on academic performance and progression data. (Achieving the Dream et al., 2015a)

This is part of a broader "guided pathways" initiative for more structure in community college education that is meant to encourage degree attainment by helping students define their end goals and generate a meaningful, highly specific route to reaching those goals; by monitoring their progress and providing feedback and resources along the way; and by providing on-ramps (e.g., introductory courses in their chosen field of interest, success skills embedded in credit-bearing courses, and accelerated developmental course sequences for students with developmental education needs) (Jenkins & Sung-Woo, 2014; Johnstone, 2015).

Texas has been the leader in the reform of mathematics remediation. One Texas project that has gained substantial traction in the arena of community college developmental mathematics reform is The New Mathways Project (NMP) (Charles A. Dana Center, n.d.). The Charles A. Dana Center at The University of Texas at Austin and the Texas Association of Community Colleges, which represents Texas' 50 independent community colleges, have joined forces to develop and implement NMP statewide. Over one-third of Texas community college systems implemented the NMP in 2014 and 47 college systems were implementing the NMP as of spring 2015[4] (Rutschow & Diamond, 2015). The NMP Transfer Champions Initiative addressed transferability of NMP college-level courses and, as of fall 2014 more than 75% of public four-year colleges and universities accepted one or more of these courses as a college-level math class for some majors (Rutschow & Diamond, 2015). The projects' main aspiration is to reclaim the mathematical lives of students who repeatedly fail algebra-based courses irrelevant to their program of study (and unnecessary for approximately 95% of professions (Rutschow & Diamond, 2015)). NMP's systemic approach to reform is built around three mathematics pathways (Dana Center Statistics Pathway, Dana Center Quantitative Literacy Pathway, Dana Center STEM Pathway) that move students through the developmental sequence faster than traditional approaches, and a supporting student

[4] In college systems with more than one institution, at least one campus was implementing the NMP.

success course (Frameworks for Mathematics and Collegiate Learning). The project, which began in 2012, is based on four fundamental principles:

1. Multiple pathways with relevant and challenging mathematics content aligned to specific fields of study
2. Acceleration that allows students to complete a college-level math courses more quickly than in the traditional developmental math sequence
3. Intentional use of strategies to help students develop skills as learners
4. Curriculum design and pedagogy based on proven practice (Dorsey, Carvalho, & Stano, 2014)

NMP goals are well aligned with the six Core Principles of the Student Success Strategy that were listed above and many of these recently published principles have already been realized at colleges involved in the implementation of NMP (Treisman, 2015).

These new mathematics courses and pathways are designed not only to help students meet a formal requirement, but also develop the quantitative literacy skills necessary for the broad range of careers community college students are pursuing. NMP courses are all developed around eight research-based curriculum design standards that are critically important to the current study. The curriculum and instruction should: 1) focus on broad math concepts and their interrelationships, not just facts and skills; 2) promote active learning inside and outside of the class, where students must formulate hypotheses, test ideas, and develop metacognitive strategies; 3) help students develop the determination to productively persist[5] in mathematics through lesson scaffolding, assisting students in a reconceptualization of struggle as important to the process, and

[5] Productive persistence was coined by Professor of Mathematics and Charles A. Dana Center Director, Uri Treisman. It is elsewhere called academic perseverance and constructive perseverance.

encouraging confidence; 4) enable students to transfer their knowledge to novel problems by including tasks that have multiple approaches and solutions and allowing students time to reflect on their problem-solving strategies; 5) contextualize math in varied disciplines by using data from real-world sources; 6) stress the importance of, and address difficulties with, domain-specific terminology, language, and symbols; 7) contain authentic reading and writing components that support students' abilities to read and write about math or statistics and use math or statistics in explanations of phenomena; and 8) utilize technology for deep involvement with the concepts. (Dorsey, Carvalho, & Castillo, 2014)

All three pathways begin with a course called Foundations of Mathematical Reasoning (Foundations). A report by the MDRC[6] (Rutschow & Diamond, 2015) elucidates initial results of student success in—and after—Foundations courses at the nine NMP codevelopment colleges: Almost 65% of the 233 fall 2013 Foundations students met their developmental math requirement by passing with an A, B, or C, and 30% of those students completed a college-level statistics course in spring 2014. By comparison, of the fall 2013 students enrolled in traditional developmental math courses at the same institutions, less than 26% met their developmental math requirement and less than 9% completed a college-level math class. Note that 16% percent of students who passed Foundations in fall 2013 were enrolled in a college-level statistics course that they did not pass in spring 2014. Results from the Foundations and subsequent college-credit-bearing course are very promising when contrasted with the national trend in which approximately 80% of students do not pass a credit-bearing course within their first three years of enrollment.

---

[6] The MDRC is a nonpartisan organization that researches the impact of policies and programs on the lives of persons of low socioeconomic status.

RESEARCH QUESTIONS

My central question is whether these accelerated, alternative course pathways are realizing the hopes of their creators. In brief: Will students in these pathways develop what I call developmental mathematics noncognitive factors—the malleable, demonstrable, research-based, noncognitive factors most critical to student understanding and achievement in developmental mathematics courses? In this study, I will address particular subgoals of this work on the extent to which students in the New Mathways Project Foundations of Mathematical Reasoning course increase their math equanimity, self-efficacy, math mindset, math belongingness, and college belongingness; and I will look for evidence of transferability of their knowledge to novel problem situations. From research, we know we can train students, at least in the short term, to execute procedural algorithms, but crucial questions remain: Can we help students see math in the world around them and then apply what they learn in their courses to unfamiliar situations of relevance to their lives? Is there evidence that students in these new courses see themselves as able to productively wrestle with and solve problems in unfamiliar domains?

Because measuring noncognitive factors generally involves the use of long, time-consuming, and potentially unreliable surveys, an additional goal of my work involves examining the viability of short, retrospective self-report measures of affective constructs in math education research. This is a dissertation about practice and I want to find ways to quickly and efficiently uncover reliable information about latent constructs[7]. Researchers have noted several issues that arise from self-report measures of change; one

[7] Practical Measurement researchers at the Carnegie Foundation for the Advancement of Teaching and David Yeager at The University of Texas at Austin influenced much of my thinking in this area. Where appropriate, I received permission from David Yeager to cite his work and note the findings were preliminary.

of these is response shift[8] bias (Drennan & Hyde, 2008; Howard, 1980; Klatt & Taylor-Powell, 2005; Nimon, 2014; Norman, 2003). According to Drennan and Hyde (Drennan & Hyde, 2008), "response shift bias occurs when the student's internal frame of reference of the construct being measured…changes between the pre-test and the post-test due to the influence of the educational programme" (Drennan & Hyde, 2008). I used a retrospective pretest design combined with a pretest-posttest design, specifically a pre-post-then design, to study whether response shift has occurred. This post hoc design is not widely used in education research, but, it is growing in popularity and if the retrospective pretest successfully measures these constructs, this finding could enable stakeholders to discern educational quality when obtainment of pre-measures is not feasible (Klatt & Taylor-Powell, 2005). Additionally, because the retrospective pretest (i.e., then-survey) used in this study is very short (approximately 3-5 minutes) and is administered at only one time point, its use in the classroom could provide instructors and administrators with an uncommonly quick measure of educational quality.

Specifically, my research questions and hypotheses are:

1. Do students exhibit differences over time in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness?

   RQ1A. Do students exhibit beginning-of-semester to end-of-semester differences in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness?

   RQ1A Research Hypothesis: Students will exhibit beginning-of-semester to end-of-semester improvements in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness.

[8] "Response-shift" carries the same meaning as "response shift".

RQ1B. Do students exhibit pre-survey to then-survey differences in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness?

RQ1B Research Hypothesis: Students will exhibit pre-survey to then-survey differences in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness.

2. Do beginning-of-semester to end-of-semester differences in students' math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness predict semester outcomes (math course grade, math final exam grade, math course percent attendance, and Developmental Assessment of Place Value Understanding score)?

   RQ2 Research Hypothesis: Beginning-of-semester to end-of-semester differences in students' math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness will predict semester outcomes.

3. Do students exhibit evidence of their ability to transfer their knowledge to novel place value problems?

   RQ3 Research Hypothesis: Students will exhibit evidence of their ability to transfer their knowledge to novel place value problems.

The participants are students enrolled at two Texas community colleges that have progressed to full-scale implementation of NMP's Foundations course. I use both qualitative and quantitative measures to examine my research questions. Measures include: pre-post-then self-report surveys on math equanimity, math and college belongingness, math self-efficacy, and math mindset; a self-report demographic survey; an instructor-reported math course grade; an instructor-reported math course final exam grade; and instructor-reported attendance. In addition to the above measures, a subset of students took the online Developmental Assessment of Place Value Understanding (DAPVU) and I conducted one-on-one semi-structured interviews with a smaller subset of students. I used an exploratory approach analyzing qualitative data and analyzed the quantitative data using multilevel models.

DELIMITATIONS AND LIMITATIONS

I have limited the scope of this study to a non-exhaustive set of developmental mathematics noncognitive factors (dm-noncognitive factors): math equanimity, math mindset, math self-efficacy, and math belongingness, and college belongingness. These particular factors were chosen because they are directly or indirectly addressed in the curriculum and there is substantial evidence that they are crucial for student mathematical success and/or causal antecedents of factors that are crucial for student mathematical success. Exclusion of other noncognitive factors does not imply that they are necessarily less important and research should examine the impact of those factors on student mathematical success. I used a modified version of the Assessment of Place Value Understanding (APVU) to gauge concept transfer. The APVU was developed by Mary Hannigan (Hannigan, 1998) and Tracy Rusch (Rusch, 1997) for their dissertation studies. Because this instrument was designed and validated to measure explicit place value understanding of pre-service elementary teachers who were in classes that covered place value, I made slight modifications to reflect the difference in populations and changed some framing to better surface transferability. This threatens the validity of this instrument and may not provide an accurate assessment of the students' ability to transfer, a notoriously difficult construct to detect. However, the main changes were based on the authors' recommendations on how to better surface conceptual understanding and this coincides with my desire to uncover deep reasoning that is often unobserved in traditional assessments of transfer. I assessed several dm-noncognitive factors with self-report surveys. Self-report surveys are prone to various biases, but I attempted to account for many of these in my survey design and looked for evidence of posttest response shift bias using a pre-post-then study design. Finally, I chose to examine a particular developmental math course and, while other nontraditional developmental math courses may share common features, results cannot be generalized to those courses.

This study has several limitations. First and most weighty are threats to internal validity. The colleges in this study had moved to full-scale implementation of the curriculum and there is no means of obtaining a comparison group. As such, significant changes in the students' dm-noncognitive factors cannot support the claim this was a direct result of the curriculum (history bias). Students were unable take the pre-survey until the second or third week of class. Because the students may have been introduced to intervention-targeted concepts prior to taking the pre-survey, it cannot strictly be considered a pre-intervention measure. I conducted one-on-one semi-structured interviews to help mitigate these limitations, but the number of interviews was too small to make true assumptions about the impact of the curriculum. Another potential threat involves maturation—the post-measures were completed at the end of the semester when students were potentially tired and overwhelmed by final exams. This study is also threatened by experimental mortality because I was unable to obtain post-measures from students who dropped the course.

The APVU was originally validated as a 50-minute, paper and pencil, in-class assessment. Unfortunately, the Foundations curriculum does not allow sufficient time for in-class administration of the DAPVU. It is also problematic to send the DAPVU home with students because this would require even greater assistance by the instructors and I would be unable to monitor how much time students spent on the assessment. For these reasons, the DAPVU needed to be administered online with added prompts that ask students to explain their written work and some problems needed to be removed to counterbalance the added time commitment. I made every effort to keep the online, shortened version of the assessment as parallel to the original as possible, but these modifications threaten the validity of the DAPVU and limit the usability of the APVU rubrics.

There are also threats to external validity: This is a short-term, small-scale study using a convenience sample (two accessible institutions) with students who volunteered to participate (selection bias). The institutions in the study demonstrated strong buy-in of the curriculum, so fidelity of implementation may be distinctly different at other institutions. Because of these issues, findings may not represent the typical user and are not generalizable to the population. A larger, randomized controlled trial could more adequately assess the impact of the curriculum on students' dm-noncognitive factors.

While this is primarily an exploratory study with multiple limitations, I believe it can make valuable contributions by providing information and feedback to the creators of a developmental mathematics curriculum that is rapidly growing in Texas and elsewhere; illuminating links between affective dm-noncognitive factors, concept transfer, and course outcomes; and adding to the research on underutilized self-report measures that could minimize time and cost of data collection.

Document Roadmap

In Chapter 1, I provided the rationale behind, and a brief overview of, the current study. Chapter 2 includes a review of literature essential to this research. In Chapter 3, I describe the instruments used in this study, the methods for data collection, and my analysis methodologies. I present the results of the study in Chapter 4 and discuss possible implications and directions for future research in Chapter 5.

Chapter 2: Literature Review

In this chapter I define dm-noncognitive factors and discuss research about the dm-noncognitive factors used in this study: math equanimity, math belongingness, college belongingness, math self-efficacy, and math mindset. Next, I provide background information about transfer research and the ways in which the meaning of math concept transfer has evolved over time. Then, I discuss how The NMP's Foundations of Mathematical Reasoning curriculum attempts to foster productive persistence, dm-noncognitive factors, and math concept transfer. I end the chapter with a discussion about research that has uncovered methodological issues related to self-report questionnaires (response shift bias) and ways that have been proposed to address these issues (retrospective pretests).

Developmental Mathematics Noncognitive Factors

Historically, conversations about reform math and teaching and learning have focused predominately on the types of academic content knowledge and cognitive abilities needed for students to be successful; math knowledge was regarded as a set of facts or skills. In recent years, researchers have begun to more critically examine the impact of so-called noncognitive factors related to student mathematical success and have developed a broader interpretation—math knowledge includes a set of dispositions or views (Dweck, Walton, & Cohen, 2011; Farrington et al., 2012; Yeager & Walton, 2011); and their findings have impacted policy and the ways institutions are preparing K-16 students (Garcia, 2014; Mathematical Association of America's Committee on the Undergraduate Program in Mathematics, 2015; National Govenors Association Center for Best Practices & Council of Chief State School Officers, 2010; West et al., 2014).

Noncognitive factors have been referred to as noncognitive abilities or skills, social and emotional learning [SEL] competencies, 21st Century skills, and soft skills; and they include things such as persistence, self-control, perceptions of utility, affect, intrinsic and extrinsic motivation, attitudes, and mindsets about intelligence (Dweck et al., 2011; Farrington et al., 2012; Garcia, 2014; West et al., 2014; Yeager, Bryk, et al., 2013). Research about the influence of noncognitive factors on learning is a relatively new field and, as such, there is not an agreed upon terminology, a comprehensive list, or a well-defined set of measures (Borghans, Duckworth, Heckman, & Weel, 2008; Bryk et al., 2013; Duckworth & Yeager, 2015; Farrington et al., 2012; Yeager, Bryk, et al., 2013).

To confuse matters more, there is not an absolute distinction between noncognitive and cognitive factors because most, if not all, noncognitive factors are influenced by cognition and they support cognitive development (Borghans et al., 2008; Farrington et al., 2012). Noncognitive factors have been loosely defined as “traits or skills not captured by assessments of cognitive ability and knowledge” (West et al., 2014, p. 1). These factors are intuitively understood to be different than intelligence and content knowledge, but this distinction is unclear and too broad to be useful in research and practice (Duckworth & Yeager, 2015). Instead of trying to separate cognition from noncognitive factors, researchers have started to look at the interplay between them. For instance, social cognitive models of motivation integrate motivational and cognitive factors. According to Linnenbrink and Pintrich (2002), there are three important assumptions of social cognitive models of motivation. First, “motivation is a dynamic, multifaceted phenomenon” (p. 313). In other words, people aren’t just generally motivated or unmotivated and researchers should consider different facets of motivation, such as attributions and goals, for educational implications. Second, “motivation is *not* a stable trait of an individual, but is more situated, contextual, and domain-specific" (p. 314). A person's mathematics self-efficacy may be different from a person's science self-efficacy, and these levels of self-efficacy can change. Lastly, "students' own thoughts about their motivation and learning play a key role in mediating their engagement and subsequent achievement" (p. 314).

Mathematical productive persistence involves navigating and persevering through mathematical situations by utilizing effective strategies and adapting ineffective strategies. "Productively persistent students are fully engaged in learning and are motivated to expend effort to reach long-term, personally meaningful goals" (Charles A. Dana Center, 2013a, p. 2). Motivation within a particular domain can predict what courses students will take, the extent to which they will persist in related educational attainment, and whether or not they will be successful in that domain (Kosovich, Hulleman, Barron, & Getty, 2014). It appears to be more useful to consider what noncognitive factors predict when students will take on a challenge in a particular domain and what persisting in the face of difficulty entails than to consider the extent to which cognition is involved. In this review, I am taking a step toward clarity about a subset of some of these well-documented affective constructs that are key to performance in developmental math courses—dm-noncognitive factors.

I define dm-noncognitive factors as the malleable, demonstrable, research-based, noncognitive factors most critical to student understanding and achievement in developmental mathematics courses. While immutable factors and traits are remarkable sources of research and can be excellent predictors of student success, I am interested in practical ways curriculum and instruction can directly help students foster the attitudes and behaviors that will enable them to persist in developing their mathematical abilities and indirectly close gaps that are generally found between students of different backgrounds. DM-noncognitive factors must be demonstrable for this research to be useful for practitioners and researchers. Many dm-noncognitive factors are latent constructs and, thus, may not be observed directly. Such constructs must be demonstrated through well-defined indicators of the underlying construct (e.g., demonstrations of, or self-report of, the related covarying behaviors), and previous research should support the constructs and measures used. In order to increase specificity, dm-noncognitive factors exclude noncognitive factors related to academics in general, such as effective learning and study strategies, even though these are likewise important and positive indicators of success. DM-noncognitive factors are inherently situated and contextual and, despite the importance of how dm-noncognitive factors may affect student success in developmental mathematics courses, there remains a paucity of studies within this demographic. As such, when outlining factors crucial to this demographic and when attempting to apply studies about noncognitive factors that were done with different demographics, it is vital to keep in mind that many developmental math students are considered non-traditional (e.g., older, minorities, low socio-economic status) and developmental math students typically have had prior negative math experiences and have underperformed on difficult tasks.

It is also important to note that some dm-noncognitive factors may precede, follow, or overlap with other factors or outcomes. For instance, improving one's math mindset could promote productive persistence and productive persistence could lead to an increase in math self-efficacy. Rather than focusing on parsing out these distinctions, I will focus on how these factors work together. The majority of dm-noncognitive factors are math domain-specific; a strong research-base must support the inclusion of any noncognitive factors that are not math-domain specific. I have limited the scope of this study to five dm-noncognitive factors: math equanimity, math belongingness, college belongingness, math self-efficacy, and math mindset. In the subsequent sections of this chapter, I examine some existing research on the significance of the aforementioned dm-noncognitive factors. Because productive persistence is so strongly intertwined with dm-noncognitive factors and because it is greatly influenced by academic mindsets[9], I will integrate relevant research on productive persistence throughout.

Math Equanimity

Anxiety "is broadly defined to be a state of emotion underpinned by qualities of fear and dread" (Hembree, 1990, p. 33). Anxiety is intuitively understood by most, and has a long research history, but it remains ambiguously defined; for instance, it has been conceptualized as "a trait, a state, a stimulus, a response, a drive, and as a motive" (Endler & Kocovski, 2001, p. 232). Originally viewed as an attitudinal construct, most researchers now conceptualize anxiety as an affective construct (though it is sometimes classified as a subcomponent of attitude) (Akin & Kurbanoglu, 2011; Ma, 1999; van Aalderen- Smeets, Walma van der Molen, & Asma, 2012).

Using Hembree's (1990) broad definition, which was based on a definition by Lewis (1970), one could define math anxiety as a state of emotion underpinned by qualities of fear and dread in mathematical situations. Math anxiety has been defined elsewhere as "feelings of tension and anxiety that interfere with the manipulation of numbers and the solving of mathematical problems in a wide variety of ordinary life and academic situations" (Richardson & Suinn (1972), as cited by Betz (1978)) and as "a

[9] According to Farrington et al. (2012), "academic mindsets are beliefs, attitudes, or ways of perceiving oneself in relation to learning and intellectual work that support academic performance." Math domain-specific academic mindsets are dm-noncognitive factors.

feeling of tension, apprehension, or fear that interferes with math performance" (Ashcraft, 2002). These definitions are more precise than the one derived from Hembree's and they reflect the majority of definitions I found on math anxiety. While I believe definitions should be as precise as possible, I do not think these definitions are appropriate. First, Richardson and Suinn's definition requires that one already understand the meaning of a word that is being defined (anxiety). More importantly, both definitions go beyond defining math anxiety by including potential *consequences* of math anxiety—interference with performance. According to Hembree, anxiety is "directed toward the future" (p. 33); it is an emotional reaction to a perceived threat.

Highly math anxious people, in the moment they are confronted with a mathematical situation (e.g., a math task) or the prospect of a mathematical situation (e.g., signing up for a math course), have a heightened sense of discomfort. This future-oriented fear is sometimes accompanied with an increased heart rate, sweating, and other physiological symptoms. According to Lyons and Beilock (2012), highly math anxious people (HMAs) can experience physical pain when awaiting a math task. Lyons and Beilock conducted a study (*N=28*) in which HMAs and low math-anxious people (LMAs) performed similarly on easy math tasks as well as easy word tasks, but the HMAs did significantly worse on the difficult math tasks than on the difficult word tasks. Prior to each task, participants received a cue to let them know if the subsequent task would be a math task or a word task. Lyons and Beilock used functional magnetic resonance imaging (fMRI)[10] to measure the participants' brain neural activity before and

[10] Eklund, Nichols, and Knutsson (2016) question the results of fMRI studies conducted between 2000 and 2015 in a recently released report. I have included two fMRI studies in this dissertation with conclusions consistent with theories and conclusions from other non-fMRI studies, but the results of the fMRI studies should be interpreted with caution.

during the tasks. Results showed that being highly math anxious predicts an increase in activity in the parts of the brain that perceive pain when anticipating a math task, but not while actually completing the math task.

If math anxiety is an anticipatory emotional state, why are its potential outcomes included in so many definitions? The reason, I believe, is because math anxiety has been demonstrated to be a powerful predictor of math avoidance and failure.

This is something Hembree (1990) and Ma (1999) shed light on when they published results of their meta-analyses of math anxiety studies. Many researchers initially believed test-anxiety theory was sufficient to cover math anxiety, some viewed math anxiety as a domain-specific test anxiety, and others defined it very broadly (Hembree, 1990; Ma, 1999). There were two main theoretical models. In the interference model, math anxiety is "a disturbance of the recall of prior mathematics knowledge and experience" and being HMA leads to poor achievement (Ma, 1999). In the deficits model, poor study and test-taking skills, not anxiety, leads to poor math performance; and recall of prior math failures leads to a person being HMA (Ma, 1999). At the time Hembree conducted his meta-analysis, the field did not yet have a solid theoretical base for the math anxiety construct (Betz, 1978; Hembree, 1990).

Hembree (1990) established that "math anxiety depresses performance" (p. 44); math anxiety causes people to fail at math or avoid math courses and math-related careers. Hembree's conclusion applies mostly to college students because the majority of the studies included in his analysis were conducted with college students. However, the negative correlation between math anxiety and math performance and attainments has been repeatedly demonstrated across many groups (e.g., age groups, gender groups, ethnic groups) and across measurement instruments (Ashcraft, 2002; Ashcraft & Kirk, 2001; Ashcraft & Krause, 2007; Beilock, Gunderson, Ramirez, & Levine, 2010; Ma, 1999; Maloney & Beilock, 2012; Young, Wu, & Menon, 2012). In other words, regardless of whether you are, say, male or female, if you have the same level of math anxiety with regards to a particular math task, this will likely impact your performance in the same way and your level of math anxiety would be relatively consistent across multiple math anxiety scales (e.g., Mathematics Anxiety Rating Scale, Fennema-Sherman Mathematics Attitude Scales). Ma (1999) also found that the strength of relationship between math anxiety and performance was dependent on the performance assessment measure, with standardized achievement tests exhibiting a weaker relationship than math school grades and instruments created by researchers. Due to the situational nature of math anxiety, however, the negative correlation between math anxiety and math performance is not consistent across different types of math (e.g., arithmetic, algebra) (Ashcraft & Krause, 2007).

As is the case with most research, not all researchers will come to the same conclusions. Arousal theorists may view Hembree's (1990) broad assertion that "math anxiety depresses performance" (p. 44) as an oversimplification because the degree of math anxiety would play a role in the effect of anxiety on performance; specifically, "arousal theory indicates that some anxiety is beneficial to performance" (Ma, 1999), though there is a point at which it becomes disadvantageous. Ma (1999) discusses three outlying studies from her meta-analysis that contradict Hembree's broad conclusion. Two of the studies found a positive relationship between math anxiety and math performance of gifted students and one found that a decrease in math anxiety of college students with strong math backgrounds did not lead to an increase in math performance. Ma infers that there are certain subgroups (high performers in this case) for which decreases in math anxiety may not be beneficial and may even be detrimental.

Prior to Hembree's meta-analysis, some important findings had already surfaced and have since been replicated and extended—math anxiety is widespread in college and college students who took less math in high school have higher levels of math anxiety than their college counterparts (Hembree, 1990); elementary education majors report higher levels of math anxiety than any other major (Beilock et al., 2010) and they may have the same level of anxiety as developmental mathematics students (Zientek, Yetkiner, & Thompson, 2010); HMA elementary teachers pass negative attitudes, including math gender stereotypes, to students (Beilock et al., 2010; Gunderson, Ramirez, Levine, & Beilock, 2012; Maloney & Beilock, 2012); math anxiety is more prevalent with students in developmental math courses than those in non-developmental courses[11] (Hembree, 1990; Zientek et al., 2010); math anxiety is more pronounced in female developmental math students than male developmental math students (Beilock et al., 2010; Hembree, 1990; Hyde, Fennema, Ryan, Frost, & Hopp, 1990; Ma, 1999); scores on high-stakes achievement tests may severely underestimate HMAs math abilities (Ashcraft & Krause, 2007; Ramirez & Beilock, 2011); there is a strong negative correlation between math anxiety and some motivational variables (Ashcraft, 2002); behavioral and cognitive-behavioral interventions with HMAs can decrease math anxiety and restore math performance to levels similar to LMAs (Hembree, 1990).

One hypothesis about the relationship between math anxiety and performance is that HMAs are simply less competent at math. Hembree's analysis showed that math

[11] Zientek, Yetkiner, and Thompson (2010) postulated that community college students enrolled in College Algebra, a credit-bearing course, might experience math anxiety similar to that experienced by developmental math students.

anxiety was significantly negatively correlated with IQ and ability, but he notes that the relationship was small. Researchers have provided ample evidence to refute the idea that math anxiety is completely confounded with math incompetence. For instance, Faust, Ashcraft, and Fleck (1992) (as described by Ashcraft and Kirk (2001)) showed that something as simple as modifying a response-time (RT) math task to an untimed, pencil-and-paper format could change outcome differences between LMAs and HMAs from significant to insignificant. Some assessments can make HMAs appear less capable than they actually are. Ramirez and Beilock (2011) used writing tasks with college students to examine the interplay between test anxiety and performance. Two groups of students were given a math pretest and then provided with an anxiety-eliciting scenario. Next, both groups wrote for 10 minutes; the expressive writing group wrote about their "thoughts and feelings" in relation to the math tasks they were about to complete and the unrelated writing group wrote about "an unrelated emotional event" (p. 212). The groups performed similarly on the pretest. However, the expressive writing group had a significant 4% pretest-posttest gain in accuracy and the unrelated writing group had a significant 7% pretest-posttest drop in accuracy. According to the researchers, "[w]riting about negative thoughts and worries accounts for choking-under-pressure differences across unrelated and expressive writing groups" (p. 212). This study was about test anxiety, but research has produced similar outcomes for math anxiety (Maloney & Beilock, 2012).

Young, Wu, and Menon (2012) used fMRI to study the brain activity of 46 second and third graders while they solved simple addition and subtraction problems. The LMAs and HMAs were similar in their IQ, working memory, reading, and trait anxiety, but when they were solving the math problems the HMAs showed less activity in the areas of the brain critical to working memory and attention (dorsolateral prefrontal cortex, pre-supplementary motor area, basal ganglia), as well as the areas crucial for math cognition (posterior parietal cortex). LMAs used areas of the brain that help with performing tasks efficiently, while the HMAs used areas of the brain that regulate undesirable emotions.

Working memory is "a short-term memory system that maintains, in an active state, a limited amount of information with immediate relevance to the task at hand while preventing distractions from the environment and irrelevant thoughts" (Beilock & Carr, 2005, p. 101). Math anxiety can reduce working memory, and this results in diminished performance (Ashcraft & Kirk, 2001; Ashcraft & Krause, 2007; Beilock & Carr, 2005; Faust et al., 1996; Ramirez, Gunderson, Levine, & Beilock, 2013). Because not every math situation will arouse anxiety in a math anxious person, Ashcraft and Krause (2007) propose that anxiety must be aroused for working memory to be affected. If it is aroused, the affected person will exhibit an "on-line effect"[12] of math anxiety while attending to the math task.

Further evidence that poor performance of HMAs cannot be attributed solely to math incompetence comes from Humbree's (1990) meta-analysis studies that included either a behavioral or a cognitive-behavioral math anxiety treatment. Behavioral treatments attempted to "relieve 'emotionality' toward mathematics (feelings of dread and nervous reactions)" and "cognitive-behavioral treatments attended to the worry factor[13] but also provided elements to reduce emotionality" (Hembree, 1990, p. 42). These successful treatments significantly decreased anxiety and this led to significant

[12] An "*on-line effect* on an individual's math performance [is] an effect on underlying cognitive processes as the individual performs a math task" (Ashcraft & Kirk, 2001, p. 224).

[13] At the time of the meta-analysis most measures and interventions related to math anxiety were built around the test anxiety construct. According to Liebert and Morris' (1967) formulation, the test anxiety construct is composed of a behavioral, emotionality factor and a cognitive, worry factor.

increases in performance scores to a level approaching those of LMAs. The relevant point is made by Ashcraft and Kirk (2001): “Because the treatments did not involve instruction or practice in mathematics, it is quite improbable that the treatment itself improved individuals’ math competence” (p. 225).

One of Hembree’s (1990) goals was to determine if the research supported the idea that test anxiety incorporates math anxiety. He concluded that test anxiety and math anxiety are correlated, learned, behavioral conditions. He lists the following parallels between the two constructs:

1. Mathematics and test anxieties both relate to general anxiety.
2. The differences in anxiety level regarding student ability, gender, and ethnicity are similar for both constructs.
3. Both forms affect performance in a similar fashion.
4. The constructs respond to the same treatment modes, with best relief from behavioral-related methods and little result from cognitive treatment, group counseling.
5. Improved performance relates to the relief of both constructs. (p. 44)

Despite these parallels, Hembree (1990) asserts, “it seems unlikely that mathematics anxiety is purely restricted to testing” because “only 37% of one construct’s variance is predictable from the variance of the other…Rather, [math anxiety] appears to comprise a general fear of contact with mathematics, including classes, homework, and tests” (p. 45).

As shown above, it is crucial for students to maintain a sense of relative calmness when they encounter mathematical situations. Borrowing from Maloney and Beilock (2012), I define math anxiety as “an adverse emotional reaction to math or the prospect of doing math" (p. 404). Speaking about math anxiety in terms of dm-noncognitive factors is difficult because dm-noncognitive factors are framed in a nonnegative manner. It is awkward (and nonsensical) to say I hope to see increases in students' low math anxiety. As such, the dm-noncognitive factor counterpart to math anxiety is math equanimity—a calm emotional reaction to math or the prospect of doing math. While I admit there may be far better definitions, I believe my definitions adequately reflect the construct of interest that ranges from a nonnegative emotional state to a progressively more intense negative emotional state. They also mirror the scale points in commonly used math anxiety measures where math anxiety is measured on a continuum from "not at all anxious" to "extremely anxious".

Math and College Belongingness

Sense of belonging in mathematics "involves one's personal belief that one is an accepted member of [the mathematics] community whose presence and contributions are valued" (Good, Rattan, & Dweck, 2012, p. 701). Your mathematics belongingness pertains to the degree to which you believe you fit in with the math community. If you have a strong sense of math belongingness you are not on the periphery; rather, you feel as though you are a respected, active, engaged member. A person at the opposite end of the spectrum who questions whether or not he or she belongs to a particular community is said to have belonging uncertainty. A person's level of belongingness may vary from context to context (e.g., social belongingness differs from math belongingness, though these may be related). "Individuals both high and low in general belongingness needs may be equally vulnerable to the potential negative consequences of a low sense of belonging to an academic domain" (Good et al., 2012, p. 701).

“Stereotype threat is being at risk of confirming, a self-characteristic, a negative stereotype about one’s group” and, when aroused, it can be detrimental to academic performance (Steele & Aronson, 1995, p.797). Stereotype threat may be experienced by any member of a group that has been negatively stereotyped, regardless of whether or not the group member believes the relevant stereotype (Steele, 1997). For example, a math-identified woman may be subjected to the stereotype that women do not belong in math-related careers and a threat is that she will fail in a mathematical pursuit and confirm this stereotype. Her lack of acceptance in the mathematical community could decrease her mathematical interest and derail her persistence in mathematics. Community college classes normally consist of a wide age-range of students, including some just out of high school and adult learners who have been out of school for quite some time. Research has shown that adult, non-traditional learners may be prone to stereotype threat in their math courses (hence, belonging uncertainty) because they feel underprepared and may perceive themselves as having an intellectual capacity inferior to that of younger students (Jameson & Fusco, 2014). According to Yeager, Walton, and Cohen (2013):

> Students from historically marginalized groups, like black and Latino students or women in quantitative fields, may worry more about belonging. When students worry about belonging and something goes wrong[…]it can seem like proof that they don’t belong. This can increase stress and undermine students’ motivation and engagement over time. (p. 63)

In their article, “Why do women opt out?” Good, Rattan, and Dweck (2012) showed that, at a highly selective university, regardless of a students’ gender, math belongingness can predict persistence on a math track (*N=133*). Strong sense of math belonging negatively correlated with anxiety and positively correlated with math confidence and math utility. The researchers also reported results from a longitudinal study with calculus students (*N=1005*). Their findings further elucidate the interrelationships between dm-noncognitive factors and environmental factors: perceived gender stereotypes coupled with the belief that math ability is a relatively stable trait diminished females' math belongingness and, hence, their grades and willingness to persist in math. However, females who adopted a growth mindset (discussed below) were able to maintain a strong sense of math belonging because the belief that intelligence is malleable can offset the destructiveness of gender stereotypes.

Academic belongingness is crucial for persistence. According to Yeager, Bryk, Hausman, Muhich, and Morales (2013), "when students draw early conclusions that they cannot do the work or that they do not belong, they may withhold the effort that is required to have success in the long term, which starts a negative recursive cycle that ends in either course withdrawal or failure" (p. 30). In a study with students enrolled in The New Mathways Project courses, Statway and Quantway, a single self-report survey question of belongingness[14] that was posed in the fourth week of class was the strongest predictor of course completion and, for students who remained in the course, it also strongly predicted whether or not they would receive a sufficient grade for enrollment access to the following math course (Yeager, Bryk, et al., 2013). According to Yeager et al., this astonishing discovery has been replicated across colleges with a large number of students in multiple studies.

A strong sense of connectedness to a math community can improve performance outcomes, but, as demonstrated by Walton, Cohen, Cwir, and Spencer (2012), even

[14] Yeager and others have separately demonstrated how a single math belongingness survey question and a single college belongingness survey question can each be a powerful predictor of developmental math course success, and I am using the same survey items in the present study.

"*mere belonging*—small cues of social connectedness to another person or group in a performance domain"—can improve achievement motivation (p. 529). In one of their four mere belonging studies (experiment 2 on p. 519), the researchers gave undergraduates a report to read that was allegedly written by a math department graduate. In one condition, the supposed graduate was listed as having the same birthday ("a trivial but identity-relevant attribute") as the undergraduate; in the other condition, the undergraduate and the graduate had different birthdays. Students in the same-birthday condition persisted longer on a domain-relevant task (an insolvable math puzzle) and reported greater math motivation and social connectedness to math than students in the different-birthday condition.

In a longitudinal study, Hausmann, Schofield, and Woods (2007) demonstrated that improving students' subjective institutional belongingness, by having university administrators send written correspondence to students and giving students university-related items (e.g., decals with the university's logo), can lead to greater student persistence in academic endeavors. There was a decline in the students' sense of belonging over the semester, but the researchers hypothesized this was due to the "newness" of college wearing off. The researchers also noted that support from peers and parents may play an especially important role in the fostering of institutional belongingness in African American students at universities composed of mostly White students.

Walton and Cohen (2011) conducted a randomized controlled trial ($N=92$) with Black and White college freshmen in which they utilized a one-hour, social-psychological treatment to alter the way students would interpret difficulties in their first year of college. The participants read findings and select quotes from a fabricated survey of upperclassmen and were told the survey findings were representative of all racial and gender groups. Then the participants wrote an essay and gave a speech about the ways things change over time in college. The materials in the treatment group focused on how the upperclassmen worried about not fitting in at first and how they made friends over time, while the control group's materials concentrated on how the upperclassmen developed more refined social-political attitudes over time.

Based on multiple measures (self-report surveys, videos, daily diaries, GPA), they found that the White students were mostly unaffected in both conditions. Black students in the treatment condition maintained higher levels of belongingness on difficult days and self-reported more achievement behaviors than Black students in the control group. The White students showed no differences in achievement outcomes the following semester. However, Black students in the intervention group achieved grades one-third of a grade point higher than the Black students in the control group, as well as all other non-participating Black students on campus. The most astounding result comes from a follow up analysis—Walton and Cohen tracked students' grades through their senior year and, "[o]verall, the social belonging intervention administered in the spring of students' freshmen year reduced the White-Black gap in raw GPA from sophomore-through-senior year by 52%" (Walton & Carr, 2011, p. 19).

Minorities, females, and small number groups may view difficulties as specific to their demographic and as proof that they don't belong. "Such interpretations undermine health, well-being, and academic performance" (Brummelman & Walton, 2015, p. 24) and just a simple reframing of academic struggle as applicable to all groups can lead to greater persistence and, ultimately, academic success.

Math Self-Efficacy

Perceived self-efficacy is a type of personal expectancy that was defined by social cognitive theorist Albert Bandura (1997) as the “beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments” (p. 3). Self-efficacy beliefs are future-oriented, situational, and domain-specific (Goddard, Hoy, & Hoy, 2004). Context may dictate and change a particular level of self-efficacy and, because self-efficacy is changeable, this construct is generally studied during instructional interventions and in terms of baseline individual differences. Bandura’s (1977) foundational article on self-efficacy initiated wide-ranging studies that have expanded researchers’ understanding of the roles self-efficacy plays in diverse domains, and self-efficacy has been shown to be a predictor of a broad variety of outcomes, such as academic achievements, reductions in anxiety, teachers’ instructional practices, career choice, social skills, and smoking cessation (Klatt & Taylor-Powell, 2005; Pajares, 1996; Schunk, 1991; Zimmerman, 2000).

Bandura (2001) situated his concept of self-efficacy within a social cognitive theory (SCT) of human behavior, where human agency, with efficacy beliefs at its foundation, is the vehicle of change. One assumption of SCT involves the “triadic reciprocality” of psychological functioning in which behavioral, cognitive, and environmental factors influence one another in a bidirectional, reciprocal fashion (Bandura, 1978; Denler, Wolters, & Benzon, 2014; Pajares, 1996). Pajares (1996) claims that “Bandura’s conception of *reciprocal determinism*” is rooted in the idea that a person’s environments and self-beliefs are informed and changed by his or her perceptions of past performances, and these revised environments and self-beliefs inform and change future performances (p. 544).[15] A person's confidence in his or her ability to successfully master a situation has a direct impact on how he or she acts in that situation. These beliefs can sometimes be better predictors of performance than predictions based on past performances (Pajares, 1997; Pajares & Miller, 1994). According to self-efficacy theory, the more efficacious a person is about a given activity, the more likely he or she will choose to participate in that activity and show persistence and resilience in the face of difficulty (Pajares, 1996).

According to self-efficacy theory, four types of experience influence a person's efficacy appraisal: enactive attainment, vicarious experience, forms of persuasion, and physiological indices (Schunk, 1991; Usher & Pajares, 2009; Zimmerman, 2000). Enactive attainments, or performance accomplishments, have the strongest impact on a person's efficacy, with successes increasing efficacy and failures decreasing it (Schunk, 1991). Mastery experiences are most influential when they are the result of prevailing in the face of difficulty, especially if the undertaking is difficult for similar others (Usher & Pajares, 2009). Notably, the stronger a person's sense of efficacy is, the less susceptible it is to malleability when the person is confronted with a failure (Schunk, 1991; Usher & Pajares, 2009). Vicarious experiences come from observations of others. Comparing oneself to a similar other working on a task provides the vicarious information needed to appraise efficacy. It is important for the model to be viewed as similar; otherwise, the vicarious experience may be ignored as irrelevant (Schunk, 1991; Usher & Pajares, 2009; Zimmerman, 2000). Efficacy information attained vicariously is more susceptible to change from future failures and has less of an impact on efficacy than direct feedback from one's own performances (Schunk, 1991). Authentic social persuasion can positively

[15] Refer to Bandura (1978) for a lengthy discussion on reciprocal determinism.

impact efficacy (e.g., "You can figure this out"), especially when it comes from a reliable source (Zimmerman, 2000), but permanence of its effect is strongly related to whether or not future attempts are successful (Schunk, 1991). Social persuasion is less influential on a person's self-efficacy than enactive attainment and vicarious experiences. Usher and Pajares (2009) caution that "it may actually be easier to undermine an individual's self-efficacy through social persuasions than to enhance it" (p. 90). Persons also take into consideration emotional and physiological cues when appraising their self-efficacy. This could include fatigue, their current stress level, or some other physical manifestation of an emotion that could be perceived as an indicator of ineptitude. In a given scenario, these four sources of information—enactive attainment, vicarious experience, persuasion, and physiological cues—are taken together to provide a person with an overall cognitive appraisal of his or her self-efficacy in that scenario.

Self-efficacy has been conflated with various related constructs, making research in this area lead to different results. There has been extensive overlap and conflicting accounts of self-efficacy and other expectancy constructs. According to Pajares (1996), this is partially due to the abundance of expectancy constructs (e.g., task-specific self-concept, self-concept of ability, expectancy for success, perceptions of competence, self-perceptions of ability, perceived ability, self-appraisals of ability, perceived control, subjective competence, and confidence), many of which are defined and/or assessed in nearly the same manner. While all expectancy beliefs are beliefs about perceived capabilities, self-efficacy differs from other expectancy beliefs because it involves a person's assessment of his or her capability with regards to a highly specific situation (as opposed to a general, global one) and is sensitive to contextual factors, such as one's management of thought processes and motivations or external environmental conditions.

Pajares explains how the conflation of self-efficacy with other expectancy beliefs has led to inaccurate measurements of self-efficacy: Many researchers who have produced conflicting reports of self-efficacy or have reported self-efficacy as a non- or negligibly-predictive variable had operationalized their variables in such a way that their instruments actually reflected measures of something more broad or decontextualized than self-efficacy (e.g., generalized personality traits, confidence). Measures of self-efficacy must be closely linked to a specific outcome if the researcher's goal is to realize its predictive and explanatory power (Bandura, 1982; Pajares, 1996).

Perceived control involves "general expectancies about whether outcomes are controlled by one's behavior or by external forces, and it is theorized that an internal locus of control should support self-directed courses of action, whereas an external locus of control should discourage them" (Zimmerman, 2000, p. 85). Schunk (1991) used an example to compare Skinner and colleagues' model of perceived control—which contains means-end beliefs, capacity beliefs, and control beliefs—to self-efficacy to make the differences more clear and to show how self-efficacy may be seem akin to capacity beliefs. Schunk's example relates to effort: "An individual might hold a means-end belief that studying hard will produce a good grade, a capacity belief that he or she can study hard, and a control belief that he or she can get a good grade" (p. 210). The distinction between perceived control and self-efficacy is well defined in accounts where outcomes do not coincide with quality of performance. Locus of control measures differ from self-efficacy measures by their lack of task and domain specificity. According to Zimmerman (2000) locus of control scales have questionable utility and have been shown to be inferior to measures of self-efficacy in anxiety studies.

Self-efficacy theory is distinguished from expectancy-value theory (where people base their behaviors on their perceived chances of meeting particular goals and how much they value those goals) because self-efficacy strictly focuses on students' perceptions of their abilities to learn and apply what they have learned without the additional caveat of whether or not they value the outcome (Eccles & Wigfield, 1995; Schunk, 1991). Bandura (1977) claims that expectancy-value theorists have mostly focused on outcome expectations, even though efficacy expectations have greater predictive power of performance and choice. Zimmerman (2000) provides an example of a study by Shell, Murphy and Bruning (1989) that measured people's perceptions about their capabilities with regards to reading and writing activities (self-efficacy) and the value people thought these activities would have for achieving various pursuits (outcome expectancies). Together, both constructs predicted 32% of the variance in reading achievement, with perceived efficacy accounting for the majority of it. Perceived self-efficacy predicted writing achievement, but outcome expectancies did not. Expectancy-value theorists Wigfield and Eccles (2000) agree with the assertion that efficacy expectations are stronger predictors, but that the expectancy construct used in their studies more closely resembles Bandura's efficacy expectation than outcome expectations because they focus on peoples' own expectations for success, rather than an outcome expectancy.

Attributions are also related to self-efficacy in that they serve as a cue for efficacy appraisal (Schunk, 1991). For example, students may attribute positive and negative performances to aptitude, amount of energy expended, perceived task difficulty, and luck. Students with a fixed math mindset may attribute a failure in math to their immutable, innate intelligence; this attribution will likewise influence their efficacy expectancies of future scenarios and indirectly influence subsequent performance.

Self-efficacy is differentiated from self-concept. Self-concept is more general and includes many forms of self-knowledge and self-interpretive feelings. According to Schunk (1991), “self-concept is an individual’s collective self-perceptions that are (a) formed through experiences with, and interpretations of, the environment and (b) heavily influenced by reinforcements and evaluations by significant other persons” (p. 211). Schunk describes self-concept as encompassing self-esteem (respect and acceptance of oneself), self-confidence (belief in one’s ability to perform competently), stability (facility or difficulty of changing one’s self-concept), and self-crystalization (degree of structure to one’s self-beliefs). Stability varies based on the degree to which a person’s beliefs are set and repeated comparable experiences help crystalize beliefs. Schunk also claims that self-efficacy is included at a low level[16] in the hierarchy of the global self-concept due to its domain-specificity. For instance, persons may possess a high academic self-concept, but self-efficacy in algebra may differ from self-efficacy in geometry, and may differ even further in reading. Domain-specific measures of different self-concept constructs help elucidate differences between these constructs. For example (from Zimmerman (2000, p. 84)), a self-esteem question could be: “How good are you at English?”, while a self-efficacy item could ask: “How certain are you that you can diagram this sentence?” The former focuses on self-esteem reactions and the latter focuses on task-specific performance expectations. While many domain-specific self-concepts are correlates of self-efficacy, well-defined self-efficacy measures have been shown to have predictive advantages.

[16] Low level does not imply lesser importance; it is merely a more specific, subarea of self-concept.

Mathematics self-efficacy may be defined in terms of beliefs in one's capabilities to complete particular mathematical tasks or meet math course objectives. It plays a predictive and/or meditational role with a wide range of variables, including students' persistence, effort, emotional reactions (e.g., anxiety), activity choices, and achievement outcomes. Pajares and Miller (1994) conducted a path analysis study with 350 undergraduates to determine the direct and indirect effects between math self-concept, math-anxiety, math perceived usefulness, gender, and extent of high school and college math. They found that math self-efficacy predicted problem solving better than math self-concept, perceived usefulness, prior math experiences, and gender, and it served as a mediator of the effect of gender and prior experience on math self-concept, perceived usefulness, and math performance. Inclusion of self-concept and self-efficacy in their regression model provided evidence that self-efficacy beliefs could demonstrate discriminant validity by independently predicting subsequent math outcomes. They also demonstrated that the college females had lower self-efficacy than males.

Schunk (1981) showed that increasing students' self-efficacy through an instructional intervention can improve math performance on division problems and increase math problem-solving productive persistence with students with very low arithmetic achievement, persistence, and confidence. Notably, the control students in Schunk's study showed no significant changes in their math self-efficacy and actually became even less persistent at solving the problems. Students who were more math self-efficacious post-intervention in Bandura and Schunk's (1981) goal-setting study also showed greater perseverance; the students in this study likewise had prior negative math experiences and began the study with low math interest.

A student's perceptions about his or her capabilities to cognitively process information can impact motivation and learning. According to Pintrich and De Groot (1990), "students who believe they are capable engage in more metacognition, use more cognitive strategies, and are more likely to persist at a task than students who do not believe the can perform the task" (p. 34). Zimmerman (2008) demonstrated that self-regulatory training can increase math self-efficacy and students' math self-efficacy beliefs are linked to an increase in math skills. In Pintrich and De Groot's (1990) correlational study with 173 seventh graders, self-efficacy was positively related to strategy use, cognitive engagement, and performance, and "students who believed they were capable were more likely to…persist more often at difficult or uninteresting academic tasks" (p. 37). However, since self-regulation was the strongest predictor of academic performance, the researchers assert "students need to have both the 'will' and the 'skill' to be successful in classrooms" (p. 38). The above findings and a large number of other self-efficacy studies support the claims that students' math self-efficacy beliefs play important, distinctive roles in student motivation and achievement and are relevant to students of all backgrounds.

Instructional interventions can increase math self-efficacy to varying degrees in people who have experienced profound math failures. In Schunk's (1981) study discussed above, students with very low arithmetic achievement, persistence, and confidence were placed in a control condition or one of four treatment conditions (modeling—attribution, modeling—no attribution, didactic—attribution, didactic—no attribution). In the instructional phase, students in the modeling conditions watched and listened to an adult as he or she modeled how to solve division problems. Students were given a packet that included written explanations. Students in the didactic conditions individually studied the packet without the modeling component. In the practice phase, students in the modeling conditions received corrective modeling feedback and were referred to the relevant sections of the packet, while students in the didactic condition were only referred to relevant sections in the packet. Trainers verbally attributed successes and failures to effort for students in attribution conditions. The potential for effective instructional interventions was demonstrated as only students in the treatment conditions developed enhanced self-efficacy.

Teacher and peer models can provide vicarious information that is needed to appraise self-efficacy. Schunk (1991) discusses a study by Brown and Inouye (1978) in which college students' performances on anagram tasks were compared to the performance of a model (as the same or better than the model). After watching the model fail, students who had been told they were better than the model had higher efficacy and showed greater persistence when they re-attempted the anagram tasks than the other students. Instructional practices that attribute prior achievements to effort increase motivation, self-efficacy, and skill more so than stressing the benefits of effort, devoid of connections between the student's hard work and enactive attainment. Schunk provides an example to contrast these two forms of effort feedback: Greater efficacy is elicited through saying, "you've been working hard" than "you need to work hard" (p. 218).

Bandura and Schunk (1981) framed their goal-setting study with children with poor arithmetic skills and low interest in math activities as a project that would help the researchers understand how arithmetic skills develop and explained it was being done in multiple schools. One treatment group was asked to set proximal goals and another was asked to set distal goals; the former group increased their self-efficacy and developed their skills better than the latter group. This result could be because the actualization of proximal goals provided performance feedback that the students were improving their skills (Zimmerman, 2000).

Since self-efficacy beliefs are malleable and can have a profound impact on students with prior negative math experiences, curriculum and instruction should foster the development of mathematics self-efficacy by including mastery experiences, utilizing cooperative groups for vicarious experiences, helping students self-monitor, praising efforts that led to mastery experiences, and incorporating the development of other dm-noncognitive factors.

Math Mindset

Recently, policy briefs and educational reforms have paid a lot of attention to theories of intelligence and the buzzword "mindsets" has entered mainstream media, yielding both positive and negative reactions (Bryk et al., 2013; Dweck et al., 2011; Kohn, 2015; West et al., 2014; Yeager, Paunesku, Walton, & Dweck, 2013). Dweck and Leggett (1988) discuss the details of previous studies that impacted their development of a model of theories of intelligence and mindsets in "A Social-Cognitive Approach to Motivation and Personality." I will summarize these findings and then discuss the more recent mindset research.

Dweck and Leggett (1988) first discuss studies in which Diener, Dweck, and Reppucci found that students with equivalent initial math abilities approached difficult problems in two substantially different ways: Students either exhibited a "maladaptive 'helpless' response" pattern or an adaptive 'mastery-oriented' response" pattern, and each of these pattern types were characterized by different cognitions, affect, and behaviors. In studies conducted by Diener and Dweck, a group of late grade-school age students worked on a task in which they were given eight problems that they solved successfully and then four difficult problems that they did not solve. In order to study the ways students transitioned from the easier problems to the more difficult problems, the researchers allowed students to think aloud about any topic they wished after successfully completing the sixth problem, monitored changes in children's hypothesis-testing strategies, and took specific measures of students' predicted performance prior to and following failure.

Students had the same cognition-affect-behavior patterns on the problems where they were successful, but different patterns when they were confronted with the problems where they were unsuccessful. Students who sought challenges, used effective strategies, monitored their own progress, showed positive affect, and persisted were categorized as mastery-oriented. Approximately two-thirds of these students commented on how they would be able to complete the challenging tasks. In contrast, students who did not persist, perceived their struggles as evidence of low cognitive ability, moved to detrimental strategies, and showed negative affect were said to be exhibiting helpless patterns of behavior. More than two-thirds of these students discussed topics that would draw attention away from the task at hand, sometimes discussing things that would make them appear successful in other domains or non-academic pursuits. The researchers also cite subsequent studies with adults that yielded similar results.

Subsequent studies showed that these two distinct response patterns stemmed from the students' goals, where students who exhibited what the researchers call a "performance goal" were those who were most concerned with proving their ability to others and students who exhibited a "learning goal" were most interested in improving their ability. Students with a performance goal were more vulnerable to adopting maladaptive response patterns. Perceived helplessness can lead to negative affect (reducing self-esteem, increasing anxiety and shame) and task avoidance. Students with a learning goal were more likely to exhibit mastery-oriented patterns. They are more likely to seek challenges, and continued persistence working on a difficult problem has been shown to encourage positive affect. Note that it cannot be concluded that performance goals are necessarily negative; it is when performance goals outweigh learning goals that students are maladaptive and fail to persist. Further, when students with performance goals are confident in their ability, they may display mastery-oriented behaviors that make them successful. On the other hand, high confidence does not necessarily need to be coupled with a learning goal orientation to make students with a learning goal to be mastery-oriented.

Dweck and Leggett (1988) created a theory of intelligence model to classify individuals as either "incremental theorists" or "entity theorists" based on the individuals' different self-conceptualizations. Students with a performance goal question whether or not their ability is sufficient and attribute failure to low ability. If a student is an entity theorist, he or she believes intelligence is a fixed, stable trait, and is more likely to have a performance goal orientation. An entity theorist conceptualizes himself or herself "as a collection of fixed traits that can be measured and evaluated" and self-esteem is influenced by success or failure on performance goals (p. 266). Since students with a learning goal are more concerned with finding an appropriate path to success, they can attribute their failure to an unproductive approach to a problem and change their strategy. If a student is an incremental theorist, he or she believes one's intelligence can be changed through experiences and effort, and is more likely to have a learning goal orientation. An incremental theorist conceptualizes himself or herself "as a system of malleable qualities that is evolving over time" and self-esteem is gained from fulfilling learning goals (p. 266). Dweck and Leggett point out that there may be some individuals who have a combination of the theories and it is possible that the most suitable view is "a recognition of present differences in relative ability but an emphasis on individual growth in ability" (p. 263). The researchers do not claim that there is no such thing as innate ability or that it should be completely ignored, merely that it could be more valuable to focus on how one's intelligence can develop and mature. In more recent work, "fixed mindset" and "growth mindset" have been used to describe the beliefs of entity and incremental theorists, respectively.

One study (West et al., 2014) that compared noncognitive skills and achievement with eighth graders from 22 open-enrollment, 5 oversubscribed charter, 3 charter, and 2 exam schools found that growth mindset correlated positively with attendance; behavior; and math and English language arts test-score gains at the school and student level. The findings for the relationship between the test-score gains and other noncognitive factors (conscientiousness, self-control, grit) reversed when the researchers aggregated the data to the school level and students at oversubscribed charter schools, on average, reported lower on these measures even though they had very large test-score gains. This puzzling situation did not occur with mindset, suggesting that it may be an essential dm-noncognitive factor.

Several studies have shown the way adults interact with students can affect students' mindsets. For instance, verbal praise that focuses on ability (or the lack thereof) can lead to students developing a fixed mindset and subsequently avoid tasks that would provide productive learning experiences but could also result in failure (Kamins & Dweck, 1999; Mueller & Dweck, 1998; Pomerantz & Kempner, 2013; Yeager, Paunesku, et al., 2013).

Various interventions have shown positive results in helping students develop growth mindsets. In a voluntary study (Blackwell, Trzesniewski, & Dweck, 2007) with 91 relatively low-achieving, socieconomically disadvantaged, minority seventh graders, students were told they were participating in a workshop that would teach them about how the brain works and help them with their study skills. Information about the malleability of intelligence was only included in the experimental group. The students in the experimental group showed significant changes in their theory of intelligence (adopting a stronger growth mindset), these changes were stronger than changes in the control group, and their mindset scores were significantly greater than students in the control group. The intervention students who had a stronger fixed mindset at the beginning of the study experienced a reversal in their previously declining grade trajectory post intervention, but this reversal did not occur with the control students who began with a stronger fixed mindset—grades of students in the control group continued to fall.

Yeager, Paunesku, et al. (2013) reported on a recent study conducted by the Carnegie Foundation for the Advancement of Teaching with developmental mathematics students at 21 colleges in the United States. In the study, researchers found almost 70% of the participants had math fixed mindset beliefs. Paunesku, Yeager, and collaborators (as described by Yeager, Paunesku, et al. (2013)) administered an internet-based growth-mindset intervention at one time point with community college students in California (mostly Latino students; $N=715$) in which intervention students saw a significantly higher overall GPA increase in all academic subjects than students who also learned about the brain, but did not learn about mindsets. Yeager, Paunesku, et al. (2013) looked at a subgroup of developmental math students (n=292) to explore the impact of the intervention on course retention because typical dropout rates of developmental math classes are alarmingly high. There were significant differences in the dropout rate between students who received a growth mindset treatment and those who did not receive the treatment—only 9% of students in the intervention group did not complete the course while 20% of students in the control group dropped the course. Yeager, Paunesku, et al. concluded: “Learning that your ‘math brain’ can grow and develop could benefit even adults who have likely experienced a lifetime of feeling ‘dumb’ at math” (p. 10).

The above examples show much promise in promoting students’ growth mindsets, but this work has been met with some resistance, most notably by Alfie Kohn (2015). Kohn describes the popularity of what he calls the “‘mindset’ mindset” and explains why he believes it is detrimental to students. He pokes fun at the widespread adoption of these ideas by saying, “one half expects supporters to start referring to their smartphones as ‘effortphones’” (p. 1). His critiques are centered not on the lack of evidence—in fact, he says a long line of research supports the main ideas—but on how it is being pervasively misused. He claims that this represents a shift in educational focus to noncogntive factors and undermines the importance of paying attention to the quality of the curriculum, authenticity of assessments, instructional methods. He also cautions that praise of effort over ability may serve as a sign to students that they are incapable. Praise can also be construed as manipulative and make students pay unhealthy attention to external rewards, which could decrease their interest in the learning process. He criticizes Dweck and Miller’s studies about the positive benefits of praising effort because it did not have a condition that provided feedback without judgment or praise (e.g., “that’s an interesting viewpoint”).

Kohn (2015) discusses research by Niiya, Brook, and Crocker (2010) that showed growth mindset students who also based their self-worth on academics employed more self-handicapping behaviors to protect their self-esteem than students with a fixed mindset. His biggest issue with a focus on noncognitive skills is that effect of the social environment on "what we do and who we are" is overlooked. He relates blaming STEM field gender discrepancies on mindsets (e.g., instead of sexism) to blaming poverty on inner-city culture (e.g., instead of political factors). While Kohn recognizes some of the merit in mindset work, he emphasizes that it has its limitations—it "will get you only so far"—and we need to spend more time focusing on changing other aspects of education.

Kohn's (2015) concerns apply to noncognitive factors beyond mindsets and should not be taken lightly. I strongly believe developing students' dm-noncognitive factors is essential to their educational success. As researchers and educators strive to enhance students' educational experiences, we must keep in mind the purpose of education to provide practical learning experiences that equip students with the noncognitive *and* cognitive skills they can apply beyond formal schooling and help them attain upward mobility, and we must find ways to critically analyze whether or not our efforts are producing the desired results.

MATH CONCEPT TRANSFER

Research on transfer can be traced as far back as 1901 when Thorndike and Woodworth tested the notion that learning in a challenging domain could provide students with skills that could be adequately generalized to other domains (Lobato, 2006; Schwartz, Bransford, & Sears, 2005). It can be taken as a priori that it would be positive for students to be able to transfer what they learn in one context to another. However, what is less clear is what is meant by transfer, how much transfer is possible, how we can enable students to productively transfer, and how we can measure whether or not they will be able to do so (Bransford & Schwartz, 1999; Lobato, 2006; Packer, 2001; Schwartz et al., 2005). There remains to be an agreed upon definition of transfer and, due to this inconsistency, findings and recommendations have been mixed.

Historically, transfer has been described as "the degree to which a behavior will be repeated in a new situation" (Detterman, 1993, p. 4). Some researchers claim transfer is everywhere and can be detected, while some claim it is too difficult to find or too complicated to make it a useful area of study (Detterman, 1993; Lobato, 2006; Schwartz et al., 2005). Carraher and Schliemann (D. Carraher & Schliemann, 2002), for instance, claim transfer is a learning theory and "the metaphor underlying transfer—namely, of transporting knowledge from one concrete situation to another—is fundamentally flawed, and leads to an impoverished caricature of how learning actually works" (p.20). Some proponents of transfer research have been accused of using too broad of a definition (e.g., if you have a big enough net, you will find anything). Detterman (1993) expressed the concern that some studies that have shown far transfer have unintentionally conflated transfer with learning or merely showed that students were able to follow directions to use a particular principle or strategy. Researchers who have used a narrower definition have cited multiple examples of failed transfer, leading them to claim transfer is a useless construct or, even worse, that people appear to be incapable of showing even the most moderate ability to transfer.

Carraher and Schliemann (T. N. Carraher & Schliemann, 1985) conducted a study with Brazilian children from middle to high socioeconomic status private schools and low socioeconomic status state schools. During Piagetian interviews, they asked students to complete seven addition and four subtraction exercises and explain the methods they used. The researchers found that students preferred using a counting method over school-taught algorithms and using the school-taught algorithms led to the most incorrect answers. Many students who made symbol manipulation errors had results that were nonsensical. For instance, in subtraction problems, they may end with an answer that was larger than the two numbers given. If the desired behavior was for students to use a traditional algorithm taught in school, according to the classic definition, this is an example of failed transfer—the students who used counting methods or invented methods did not repeat the desired behavior. However, it was these students who were the most successful in solving the problems. If the definition is extended to include other, non-identical learning (e.g., learning about how to count backwards and whether or not a result is meaningful), they successfully transferred their learning.

Schwartz, Bransford, and Sears (2005) provide a similar example from research by Lave (1988) where educated adults opted to use non-traditional, invented methods when comparing prices, but did poorly on written tasks that involved the same scenarios. The subjects used situational arithmetic and, by the classic definition, did not transfer what they learned in school to the new setting. Looking for evidence of transfer, especially for far transfer, using the classic definition requires that people exactly replicate previously learned procedures to the new situation and these studies predominately have dismal results. This narrow view perpetuates the assumption that transfer is a static process and does not take into account the complexities of how people learn and seemingly capable people appear incapable of using previously learned knowledge in new ways. Several researchers[17] have made great strides in reconciling the

[17] My thinking about transfer and the format of this section has been heavily influenced by the analyses and reconceptualization of transfer by Schwartz, Bransford, and Sears and Hatanto's work in adaptive expertise. Additional supports for these views as well as different perspectives are supplied in other works

conflicting perspectives about transfer and ways to promote and assess it; I will discuss some of their findings here.

Schwartz, Bransford, and Sears (2005) recommend a reframing of educational goals and propose new ways to look at transfer that go “beyond the classic ‘stimulus generalization’ view of transfer” (p. 6). With an extension of the classical definition of transfer, examples of positive transfer could include studies that show people with greater schooling are more adept at solving novel problems in flexible, though possibly nontraditional, ways than people with less schooling (e.g., studies by Schleimann and Acioly (1989) and Lave (1989), as noted by Schwartz et al. (2005)). The authors claim that including “flexible adaptation of old responses to new settings” (p. 7) could be expanded yet further by including “preparation for future learning [PFL]” (p. 8). PFL differs from the “direct application” notion of transfer that is normally measured by “sequestered problem solving [SPS],” where persons are denied access to both feedback and additional information during the transfer situation, and “make people look much ‘dumber’…than is actually the case” (p. 9). Schwartz and Martin (2004) note that, while SPS measures serve important uses, they can be “blunt instrument[s] for assessing whether someone is ready to learn” (p. 146). Even though a common goal of education is for students to become independent learners, educational assessments are typically SPS measures that do not test for whether or not they are ready to learn.

SPS transfer assessments traditionally focus on what Broudy (1977) refers to as “replicative knowing” and “applicative knowing”, two types of knowing on which academic success is typically judged. The first type of knowing deals with recall of

---

(e.g., Beach, 1999; Bransford & Schwartz, 1999; D. Carraher & Schliemann, 2002; Detterman, 1993; Lobato, 2006; National Research Council, 2000).

specific facts or procedures, and people are notoriously unable to provide detailed accounts of what they have learned (e.g., specific dates of historical events) unless they have learned them with much repetition (e.g., basic multiplication facts). The second involves whether or not an individual directly applies certain facts, rules, and principles to determine a solution to a problem. Broudy claims that schools may be judged as failing in this aspect as well since application of knowledge (e.g., chemistry principles) to a transfer/novel context (e.g., energy crisis) is “largely confined to specialists who learn how to do it in professional education and experience on the job” (p. 10). Focusing on measuring whether or not people “know that” or “know how” leads to a wealth of examples of failed transfer. Broudy suggests a goal of schooling may also be for students to develop their “interpretive knowing” and we should consider how they are able to “know with” things they have previously learned. For example (by Broudy), we may not be able to recall facts and proofs we supposedly learned in geometry, but this prior learning provides a context for subsequent understanding of space talk.

Schwartz et al. (2005) believe that one of the main ideas behind Broudy’s interpretive knowing is that “what one notices about new situations and how one frames problems has major effects on subsequent thinking and cognitive processing” (p. 14) and this overlooked type of knowing should be integrated into the new perspectives on transfer and how we test for and promote it. Transfer studies have customarily focused only on what people “transfer out” (or, more commonly, don’t transfer out) of a learning experience, but, in their discussion of PFL, the authors suggest also considering the interpretive knowing people “transfer in” because these preconceptions can affect learning and the ways they “transfer out” of learning.

Schwartz and Martin (2004) presented results of a study with ninth-grade students using a PFL model of transfer that considered spontaneous[18] "transfer in" (comparing how two different instructional methods prepared students to learn) and spontaneous "transfer out" (comparing the double transfer paradigm with the standard transfer paradigm). There were two treatment groups: one group invented solutions for a statistics problem and the other group learned and practiced the visual procedure for solving the same problem. The researchers embedded a relevant learning resource in a subsequent test. To assess spontaneity of transfer, they did not inform the students of its relevance. Half of the students in each group were given the resource prior to attempting the transfer problem and the other half in each group had to attempt the transfer problem without the resource. This study (and their other studies with similar findings) had relatively small sample sizes so there are issues of generalization, but the researchers had some promising findings. At the posttest, there were no significant differences in the tell-and-practice with the resource, tell-and-practice without the resource, and the invention-based without the resource. However, the inventing students who received the resource provided more than double correct quantitative answers than the other conditions.

The double transfer paradigm allowed Schwartz and Martin (2004) to detect what students spontaneously transferred in, a type of knowing that would have been undetected by the standard paradigm, and the invention activities provided greater preparation for future learning, as evidenced by the inventing group's spontaneous transfer out. Another important finding that differs from traditional transfer studies is that it wasn't the type of content that influenced what students transferred in to prepare them to learn, but how the

[18]Spontaneous transfer occurs when a subject notices the analogous nature of his or her prior knowledge and the transfer situation without the analogy being highlighted or implied by the researchers. Gick and Holyoak (1983) elaborate on the difficulty of finding spontaneous transfer.

instructional activities affected their interpretive learning of the content. This supports the idea that interpretive knowing should be integrated with the replicative and applicative.

A related conception important for an expanded view of transfer is what Hatano and Inagaki (Hatano, 1988; Hatano & Inagaki, 1986) denote as “adaptive expertise”. Persons who are especially efficient at solving routine problems and executing complex predefined procedures in a particular domain are “routine experts” in that domain. They may be able to proficiently access their replicative and applicative knowing, but they are not necessarily able to apply their knowledge to novel problems in innovative ways. Adaptive experts are those who can efficiently solve common problems and can also draw on deep understandings to be flexible, consider multiple perspectives, and accurately assess their tentative knowledge when solving novel (transfer) problems (Martin, Baker Peacock, Ko, & Rudolph, 2015).

Hatano (1988) contrasts two types of experts to make the distinction clear. Studies by Amaiwa and Hatano revealed that abacus operators were highly skilled at efficiently performing multi-digit multiplication problems. While they were routine experts, they had impoverished conceptual knowledge. For example, a group of students who had been expertly using an abacus for a full year could not explain multi-digit subtraction procedures they used any better than students their same age who had just begun their abacus training. Carraher, Carraher, and Schliemann (T. N. Carraher, Carraher, & Schliemann, 2000) conducted a study with Brazilian children who were street vendors familiar with negotiating prices with customers. They found that the children were able to calculate prices for products without the use of pencil and paper, even when the underlying calculations were quite complicated. They differed from the abacus children in that they were able to draw on their solid understanding of the number system, adapting their calculation methods when necessary (e.g., inflation increases, customer needs); they were adaptive experts. Some routine experts do not develop adaptive expertise; while they may continue to hone their abilities to solve predictable problems, they may not continue to learn in other ways (Hatano & Inagaki, 1986; Hatano & Oura, 2003).

Based on further analysis of these examples, Hatano (1988) argues that there are four conditions that can help students move toward adaptive expertise: "(1) encountering novel types of problems continuously, (2) being encouraged to seek comprehension over efficiency, (3) freedom from urgent need to get external reinforcement, and (4) dialogical interaction" (p. 68). Dialogical interaction encourages students, even those lacking in prior content knowledge, to try to understand problem situations because it yields and intensifies "cognitive incongruity[19]" by helping people monitor their comprehension. Hatano's focus on these four conditions brings to light the importance of considering the sociocultural and socioemotional aspects of adaptive expertise.

As discussed above, a significant ingredient to adaptive expertise is the ability to monitor one's own preconceptions of a novel (transfer) problem in order to determine what modifications need to be made to one's approach in finding a solution. In the PFL studies above, the researchers prepared students with experiences that helped them monitor what they learned from the embedded resource and this led to successful "transfer out". If researchers are unable to create these learning experiences, SPS assessments appear to be the only option. With their focus on "knowing what" and

[19] There are three types of cognitive incongruity: (1) *surprise*, which stems from a discrepant event that shows the falsity of a preconception (2), *perplexity*, which occurs when people realize there are multiple likely explanations for a situation, and (3) *discoordination*, which involves a realization that all relevant knowledge is available, but the parts are not meaningfully connected. Notably, "for cognitive incongruity to occur, people must themselves recognize the inadequacy of their comprehension" (Hatano, 1988).

"knowing how," SPS measures do not lend themselves to detecting preconceptions; what students "know with" remains a mystery.

Schwartz et al. (2005) discuss studies in which SPS assessments that were modified to assess PFL provided glimpses into this thinking. Fifth-graders, college students, and K-12 principals were all given a challenge to develop a statewide recovery plan to increase the population of bald eagles. One study compared the fifth-graders to the college students to see if the college students would show greater transfer of their education in general than the fifth-graders and the other study focused solely of the ability of the principals to transfer from their own knowledge base. The first part of both experiments was an SPS (apply what you know) measure and all groups appeared incapable of creating workable plans. The second part of the experiment asked the students to construct questions that could provide answers which would, in turn, help them learn more about eagle recovery plans. Both groups showed greater competency and the college students' questions showed they were better prepared than the fifth graders for future learning. The second measure for the principals required them to say how they would learn to solve the problem, instead of asking them to generate questions. The principals likewise looked more competent than they appeared to be on the SPS measure and it was clear their extensive backgrounds had prepared them for future learning more so than the backgrounds of the college students. The principals also showed greater adaptive expertise in that they "resist[ed] premature assimilation" when they recognized that they might not have sufficient understanding to answer the problem. This acknowledgement, which is prerequisite for cognitive incongruity, motivated the principals to learn (Schwartz et al., 2005). In order to help themselves learn how to solve the problem, they adapted their strategies (by deciding to use the wireless network with prompting to find related information) and were making progress toward finding a solution near the end of the workshop; in other words, they productively persisted.

These studies show that, even if we are unable to provide the learning experiences like the inventing studies, we may still be able to measure what students "transfer in" and this can give us insight into what they could be able to "transfer out" if provided the appropriate experiences, even if the "transfer out" measure appears to be a case of failed transfer. It shifts the focus from deficiencies to competencies. Further, the above studies highlight the importance of finding a balance between promoting and measuring innovation and efficiency. According to Broudy (1977), "the argument for general education should be that the schooled man thinks, perceives, and judges with everything that he has studied in school, even though he cannot recall these learnings on demand" (p.12).

THE NMP'S FOUNDATIONS OF MATHEMATICAL REASONING

The New Mathways Project's Foundations of Mathematical Reasoning curriculum was designed to enable developmental mathematics students develop the attitudes and behaviors necessary to be successful in their mathematical and future career pursuits. One way it attempts to do this is by fostering students' dm-noncognitive factors (Charles A. Dana Center, 2013b). The dm-noncognitive factors in The NMP's courses feed off of each other and are included to stimulate productive persistence and transfer by emboldening students with the confidence and tenacity to press on through challenging problem situations despite the chance of failure. Some dm-noncognitive factors are introduced during the first lesson, in which students are given a "Successful Students" handout. The instructor provides a description of the handout and students work together to link policies in the syllabus that will enable them to be successful (as defined by the handout). A sample instructor's description of the handout is:

> The items listed in this handout are some mindsets and beliefs that are common among successful students. If you don't have these behaviors and mindsets yet, don't worry. This will be a semester of growth, and it is possible to develop them with a little effort. As we progress through the course, concentrate on how your beliefs, thoughts, and behaviors are affecting you, and work toward adopting these characteristics for yourself. You'll see quite an impact on your learning. (Charles A. Dana Center & Texas Association of Community Colleges, 2014)

Productive Persistence. The NMP's curricula promote productive persistence by valuing rigor; scaffolding students as they work through challenging, unfamiliar math topics; stimulating self-regulation, self-reflection, and struggle (Dorsey, Carvalho, & Castillo, 2014). The Foundations curriculum explicitly lays out levels on a continuum of productive persistence to help teachers scaffold productive struggle (Charles A. Dana Center, 2013b). At Level 1, the instructor should facilitate small- and whole-group discussions as students work through problems that are broken into sub-questions with strategy ideas. The instructor actively models how to appropriately engage in math talk and stresses that struggle is an important aspect of the process of gaining deeper understanding. There is more group work at Level 2 and strategies are not laid out. The instructor provides support to groups who need it and only moves to whole-group discussion to address misconceptions or to make conceptual links that are not immediately salient. At Level 3, problems are rarely broken into sub-questions and group-work is predominately group-directed. The instructor encourages students to continue working on difficult problems on their own prior to seeking instructor assistance. The class comes back together and groups discuss and reflect on the various solutions.

Each lesson assignment is scaffolded in a way similar to the overall curriculum (progressing from easier problems with sub-questions to more perplexing scenarios without supports) to promote entry-level successes. The curriculum provides instructors techniques for grading because there are many difficult problems that the students are unlikely to answer correctly, but these problems are included to help students "increasingly engage in productive struggle" (Charles A. Dana Center, 2013b, p. xiii). The curriculum is organized around broad concepts and spiraling for big idea connections is included through references back to previous lessons. Design Standard V, Context and Interdisciplinary Connections, also promotes productive persistence because research has shown that students in contextualized developmental mathematics courses are more apt to sign up for, and pass, credit-bearing courses.

As I described in the review of the literature, self-regulation enables productive persistence. Self-assessment involves assessing one's skills and self-regulation is when one uses that information to monitor and amend how one approaches subsequent tasks. The Developing Self-Regulation lesson (3.E) makes self-assessment and self-regulation strategies salient and helps students create an action plan for how they will self-regulate. Students are also asked to self-assess in preview assignments, rating themselves on prerequisite skills.

Math Equanimity. The Building a Learning Community (1.B and 1.D) lessons provide reasoning behind the course layout and instructor role. Students may be unacquainted with these methods and not knowing the reasoning behind these methods can lead to increases in anxiety in the math classroom. On the other hand, knowing the "why" behind classroom happenings, including why they will be given problems that may result in failure, can increase their math equanimity; prior math failures have led to detrimental outcomes and these explanations can help put their minds at ease. Early successes and strategy aids at the entry levels of productive persistence are also meant to increase their math equanimity.

Math and College Belongingness. The Building a Learning Community lessons provide an early structure for the classroom culture to be a respectful, positive environment where all students' ideas are accepted and students are situated as a community of learners working together to solve problems. Developmental mathematics students have, by definition, experienced math failures and are not accustomed to being part of a math community. Regular group work can help them reformulate their ideas about what it means to be a member of a math community and help them see that they can learn from and teach others. The instructors are also encouraged to help students realize that they care about them, since research has shown this plays a major role in student success. Suggestions include asking students to provide personal information (e.g., "tell me something interesting about yourself"), having students write a math autobiography, and establishing connection-building routines (e.g., daily greeting students as they walk into class instead of busily preparing for class). The Seeking Help lesson (3.B) makes students aware of different sources of help, including campus resources, at their disposal and communicates that help-seeking is a proactive behavior, not a sign of incompetence. It reinforces the idea that they all belong to a community striving toward similar goals.

Math Self-Efficacy. Self-efficacy is not explicitly covered in the curriculum, but it is promoted in multiple ways. Students receive preview assignments that address prerequisite skills students should master prior to the next lesson. This and entry-level successes are meant to bolster their confidence as they approach later problems. Moving through the levels of productive persistence allows students to attain mastery experiences and, since these mastery experiences will have followed their prior math failures, they could be more valuable for efficacy appraisal. Students have vicarious experiences through their regular group work; because NMP's Foundations curriculum has documented successes, it is hypothesized that a large number of these experiences will be positive. Verbal persuasion will be authentic if it comes from their similar peers. The start of each lesson is meant to be accessible to students and provide interesting discussions that could decrease their anxieties about the topic. Extraneous sources of discomfort related to physiological states may be unavoidable, but issues such as stress about college life in general could be addressed through the Seeking Help lesson.

Math Mindset. The Brain Power lesson (6.B) reinforces key elements of productive persistence, including the need to refine skills and develop understanding through deliberate practice[20]. Key concepts included in the lesson are: brain neuroplasticity, math fixed mindset, and math growth mindset. Students are asked to reflect on experiences where they succeeded at a goal after a failed attempt, and consider their levels of frustration and whether or not they tried new strategies to link these experiences to a growth view of intelligence. As students work through different productive persistence levels and gain mastery experiences, they will have growing proof that they are able to succeed in a subject in which the previously failed.

---

[20] Deliberate practice differs from repetitive practice. According to Bryk et al. (2013), "[d]eliberate practice eschews rote repetition for carefully sequenced problems developed to guide deeper understanding of core concepts" (p. 13).

Math Concept Transfer. Students in all NMP courses are required to use multiple solution methods to solve novel problems. According to the Dana Center's description of NMP's Curriculum Design Standard IV, Problem Solving, "NMP supports students in developing problem-solving skills in which they apply previously learned skills to solve nonroutine and unfamiliar problems" (Dorsey, Carvalho, & Castillo, 2014, p. 7). The organization of the curriculum around broad ideas, as opposed to discrete topics, should help students make connections between relevant topics and aid in their ability to apply what they have learned in one area to another. Math concepts are presented in authentic contexts using real data from other disciplines (Design Standard V). This is meant, not only to increase engagement, but also to help students transfer what they learn in math class to real world situations. Students explore situations involving issues such as debt and nutrition in a math context. The curriculum's focus on productive persistence can also enable students' ability to transfer because learning through struggle helps make ideas stick and ideas that are more deeply understood are more comfortable to apply in future situations (Kivel, 2014).

RESPONSE SHIFT BIAS AND RETROSPECTIVE PRETESTS (THENTESTS)

George Howard is credited with response shift bias theory and initiating a resurgence of interest in retrospective pretests (Bray, Maxwell, & Howard, 1984; Drennan & Hyde, 2008; Hill & Betz, 2005; Howard, 1980; Howard, Ralph, et al., 1979; Howard, Dailey, & Gulanick, 1979; Klatt & Taylor-Powell, 2005; Nimon, 2007). Prior to Howard's 1979 publications in which he addresses response shift bias with five studies, Campbell and Stanley (1963) provided an extensive overview of many threats to the internal validity of self-reported, pre-post measures of change (e.g., history, maturation, selection bias, instrumentation) and they claimed that these threats could best be controlled through use of true experimental designs[21]. George Howard et al. (Howard, Ralph, et al., 1979) provided experimental evidence to criticize this claim, saying that the instrumentation threat to internal validity is not always controlled through true experimental designs.

According to Campbell and Stanley (1963), instrumentation threat involves "changes in the calibration of a measuring instrument or changes in the observers or scores used" and this "may produce changes in the obtained measurements" (p. 5). Campbell and Stanley's recommendations relate to non-participant raters, but Howard et al. point out that, in the case of self-report measures, the study participants are the raters. Because participants in a treatment group have, by definition, undergone a treatment different from control group participants, their frame of reference when analyzing the meaning of questions on an instrument is potentially different. They may have a more nuanced understanding of the underlying construct than their control group counterparts; as such, they are answering questions using a different internal scale than persons in the control group and study results may be due to an instrument effect instead of a treatment effect (or lack thereof). Howard et al. refer to this threat to internal validity as "response-shift bias."[22]

Howard, Ralph, et al. (1979) documented response shift bias in five studies. In Study I, 704 non-commissioned officers at various United States Air Force bases took part in workshops designed to decrease dogmatism by making the officers more aware of

[21] Campbell and Stanley's (1963) true experimental designs are: pretest-posttest control group design, Solomon four-group design, and posttest-only control group design.

[22] Response shift bias theory is also referred to as response shift theory. Golembiewski, Billingsley, and Yeager (1976) refer to true change as "alpha" change that may be detected with a "constantly calibrated measuring instrument related to a constant conceptual domain" (p. 134). They formulate response shift bias in terms of "beta" (e.g., recalibration of intervals on a Likert scale) change and "gamma" (e.g., redefinition of concepts on a survey) change.

factors that affect communication techniques. Participants provided pretest and posttest responses on the Rokeach Dogmatism Scale (RDS), a 40-item self-report survey of dogmatism. The officers' mean posttest scores were slightly significantly higher than the mean pretest scores, as determined by a one-tailed t-test. This result did not correlate with expectations of the researchers or facilitators because they witnessed the officers become less dogmatic and it did not correlate with officers' comments on workshop evaluation forms, comments that were consistent decreased dogmatism. In follow-up interviews, many of the officers explained that they viewed their initial level of functioning differently at the end of the workshop. For example, one subject said he should have chosen "I agree very much" instead of "I disagree on the whole" on the pre-survey; he described how interacting with the workshop group members made him realize where he was initially. His ratings were -2 on the pretest and -1 on the posttest (more dogmatic), but he believed they should have been +3 on the pretest and -1 on the posttest (less dogmatic). The researchers believed that the perplexing results from quantitative analyses were due to response shift bias.

Howard and colleagues (Howard, Ralph, et al., 1979) hypothesized that using a retrospective pretest-posttest procedure instead of a traditional pretest-posttest procedure could more accurately assess changes in levels of dogmatism and eradicate potential response shift bias. While Campbell and Stanley (1963) had noted positive ways in which retrospective pretests may be added to research, Howard was first to suggest they be used to address response shift bias. In a traditional pre-post research design, the dependent variable is measured prior to the treatment and again after the treatment. A retrospective pretest, also known as a thentest, is a pretest administered at the conclusion of an intervention that asks subjects to recall how they perceived themselves or their behaviors prior to the intervention (Howard, 1980). The order and location of a thentest varies by research design—retrospective pretest designs include pre-post-then, pre-post&then, post-then, and post&then designs[23] (Nimon, Zigarmi, & Allen, 2011a). Nimon et al. (2011a, p. 12) define these four types of retrospective pretest designs as follows:

> Pre-Post-Then: Pre-post design incorporating a thentest in which the posttest and thentest are administered as two separate questionnaires, with the posttest administered before the thentest.
>
> Pre-Post&Then: Pre-post design incorporating a thentest with post and then items administered as a single questionnaire, with post items presented first.
>
> Post-Then: Posttest design incorporating a thentest in which the posttest and thentest are administered as two separate questionnaires, with the posttest administered before the thentest.
>
> Post&Then: Posttest design incorporating a thentest with post and then items administered as a single questionnaire, with post items presented first. (p. 12)

Howard, Ralph, et al. (1979) conducted a second study (*N=247*) to examine whether or not using a post&then design would yield different results than the prevalent pre-post design and signal the presence of response shift bias. The second study design mirrored the first, with the exception that officers were randomly assigned to either a pre-post condition or post&then condition[24]. Significantly more officers in the post&then condition reported decreases in dogmatism than officers in the pre-post condition. This

---

[23] This dissertation does not include all types of designs that exclude collection of pre-measures, such as designs with perceived change measures and designs substituting pre measures with research-based estimates of pretest status. See Lam and Bengo (2003) for a comparison of designs that ask participants to provide a pre-score to post-only designs that do not ask participants for a pre-score.

[24] Howard, Ralph, et al. refer to the conditions as pre/post and then/post instead of pre-post and post&then, but I am using the latter terms to remain consistent with the definitions used throughout this dissertation.

discrepancy was attributed to the differences between the thentest and the pretest because all participants took part in the workshop and posttest scores were the same for both groups. The researchers offered response shift bias theory as an explanation, but also said that these findings could be due to respondents' faulty memories or respondents' attempts to deliver positive results. The authors note that Campbell and Stanley (1963) had cautioned thentests may be susceptible to these biases. According to Campbell and Stanley, "the probable direction of memory bias is to distort the past attitudes into agreement with present ones, or into agreement with what the tenant has come to believe to be socially desirable attitudes….memory bias seems more likely to disguise rather than masquerade as a significant effect of *X* [exposure to an experimental variable or event] in these instances" (p. 72).

It is important to clarify the distinction between response shift and the theory of response shift bias. According to Nimon (2014), response shift[25] is present whenever there is "a statistically significant difference between the participants' retrospective and traditional pretest assessment of an underlying construct (e.g., self-perception of knowledge, skills, or attitudes)" (p. 258). Presence of response shift does not explicitly prove the presence of response shift bias because response shift bias theory is only one possible explanation of the underlying cognitive processes that create response shift. Two theories competing with response shift bias theory—personal recall theory and impression management theory—claim response shift is due to memory distortion (Hoogstraten, 1982; Nimon, 2014; Norman, 2003; M. Ross, 1989; Schwarz, 1999; Sprangers, 1996). Response shift bias theory purports that judgments of a pre-

[25] I solely use "response shift" when referring to a difference between pretest and thentest scores, but some researchers use "response shift" when referring to "response shift bias" and one must depend on context clues to clarifies the researchers' intended meaning. For example, see Norman (2003).

intervention state are based on changed, internal standards, whereas the theories of personal recall and impression management suggest that judgments of a pre-intervention state are based on distorted memories. Multiple studies have drawn different conclusions as to which theory is more valid and validity of a theory may be specific to each particular occasion of test (Norman, 2003). "Therefore, the argument is not for or against response shift. Instead, *the argument relates to what theory best explains the difference between participants' retrospective and traditional pretest self-assessments*" (Nimon, 2014, p. 258). Further details will be provided about the theories of personal recall and impression management in the subsequent sections.

Howard, Ralph, et al. (1979) claimed retrospective self-report measures may be more valid than prospective self-report measures due to the threat of response shift bias inherent in pre-post designs, but they remained cautious because their first two studies did not include nonsubjective outcome measures. Even if post&then designs more accurately represent respondents' perceptions of change than pre-post designs, do these perceptions correlate with authentic, nonsubjective changes? The researchers conducted further studies to analyze this question.

In Study III and Study IV, Howard and collaborators (Howard, Ralph, et al., 1979) employed a pre-post&then design in which all participants completed all data measures. Study III consisted of 51 feminine[26] women who were either assigned to the waitlist control (WL) group or randomly assigned to a discussion orientation (DO) group or a full treatment (FT) group; DO and FT groups covered the same topics and these groups were devised to promote androgyny by helping the subjects cultivate stereotypically masculine skills. Data measures included: a self-report measure of

[26] They were defined as feminine through their scores on the Bem Sex-Role Inventory.

assertiveness (CSES), an objective measure of assertiveness (OMA), a self-report measure of sex-role orientation (BSRI), and a self-report measure of change on individual goals (COI). Treatment group facilitators also rated "how much each member profited from the group experience" (p. 8) (FR). Participants completed self-report measures two months after the posttest; and completed self-report measures in post&then format and the OMA one year after the posttest.

Howard and collaborators (Howard, Ralph, et al., 1979) calculated change scores from beginning to end of intervention, beginning to two-month follow-up, and beginning to one-year follow up two times: first using the pretest scores on all self-report measures and second using thentest scores on all self-report measures. They analyzed change using one-way analyses of variance and found that post&then measures accurately represented, and traditional pre-post measures underestimated, treatment effects: Analyses using change scores with then-surveys produced significant treatment effects, while analyses using change scores with pre-surveys showed minimal treatment effects. Nonsubjective measures agreed with post&then measures more than they agreed with pre-post measures—the objective measures were more highly correlated with the post&then self-report measures than the pre-post self-report measures when looking at change from beginning to end of intervention and beginning to two-month follow-up, but there were no differences in pre-post measures and post&then measures when looking at changes from beginning of intervention to one-year follow-up.

Howard, Ralph, et al. (1979) found evidence supporting the "intuitive hypothesis that 'response-shift' effects are treatment dependent" (p. 10)—there was a wider gap between pretest and thentest mean ratings on self-report measures for treatment group participants than for control participants. Participants provided additional support to response shift bias theory when they were asked about their reactions to the thentest upon completion; persons in the control group maintained the accuracy of their pretest responses and treatment subjects detailed treatment experiences that led them to question the validity of their pretest responses. As a result, the researchers cautioned that response shift bias effects are “potential contaminants in designs employing placebo or wait-list control groups” (p. 10). Study IV was similar to Study III, both in design and conclusions drawn: quantitative analyses suggested that self-report thentests may be more accurate than self-report pretests and anecdotal accounts suggested greater concurrent validity for post&then designs than pre-post designs.

Personal Recall Theory

Personal recall theory suggests that people provide biased reports of previous behaviors and feelings based on implicit theories of stability or change. According to Ross (1989), a person’s long-term memory of a personal attribute status (e.g., feelings of math self-efficacy six months ago) may involve two steps: a benchmark assessment of his or her current attribute status and a decision about whether his or her previous state was the same or different from that benchmark state. He claims that memories that are consistent with a person’s implicit theory are more accessible than those that are not consistent and, when a person cannot recall a previous status, he or she guesses at the past status based on his or her implicit theory and benchmark status. The threats of memory bias may be greater as more time passes (Hill & Betz, 2005).

Ross (1989) reported on a previous study[27] by he and Conway (Conway & Ross, 1984) in which they demonstrated how people can find ways to support their implicit theory of change, even when no such change occurs. Participants self-assessed their study

[27] Ross did not provide details about sample size.

skills and were randomly assigned to a waitlist control group or a study skills program intervention group. All study participants self-assessed their post-intervention study skills, were asked to recall their pre-intervention ratings, and asked to guess their semester grades. The intervention did not improve grades, but the intervention group anticipated significantly higher grades than the control group, expected significantly higher grades than they actually obtained and, six months later, they believed they had received higher grades than they had obtained. Intervention participants incorrectly recalled their pre-intervention study skills ratings as much lower than the ratings actually were, while the control group did not. According to Ross (1989), the intervention group had systematic bias in grade recall. Ross claimed that the formulation and maintenance of perceived improvement made intervention participants undershoot their thentest ratings in support of their implicit theory of change.

Likewise disconcerting, in retrospective reports when a person does not recall a previous state, a person with an implicit theory of stability could demonstrate response shift with closely matching posttest and thentest scores, even if the persons underwent changes (M. Ross, 1989). Ross asserts that individuals who have undergone definite changes may even exaggerate the similarity of their current (post) and prior (then) attitudes. In personal recall theory, response shift involves a rewriting of history so that history agrees with one's implicit theory of stability or change in a given scenario. Thentests do not typically ask people to recall previous ratings like the thentest described by Ross, but his study and others showing differences between pretest ratings and remembered pretest ratings demonstrate how faulty memories could lead researchers to draw incorrect conclusions (Norman, 2003; M. Ross, 1989). Hence, many supporters of personal recall theories claim retrospective accounts are prone to more biases than prospective accounts (Taminiau-Bloem et al., 2016).

Norman (2003) contrasted the Conway and Ross (1984) study with similar studies by Sprangers and Hoogstraten (1989) to show thentest ratings can sometimes be more valid than prospective ratings when respondents hold an implicit theory of change. In Sprangers and Hoogstraten's first experiment ($N=37$), treatment subjects were in an ineffective study skills course. The intervention participants could not recall their pretest ratings on a memory test, so their thentest ratings were ostensibly based on a benchmark (post) state. The program did not have an effect (no objective post-intervention differences between control and experimental groups) and the researchers did not detect presence of response shift (no difference between pretests and thentests). In their second experiment ($N=58$), Sprangers and Hoogstraten found response shift on surveys among treatment subjects in an objectively effective training program for dental students. The students presumably believed they had changed and adjusted their thentest ratings based on this belief (as opposed to a recalibration of scale meaning), making the thentest more reliable than the pretest. According to Norman, participants in the ineffective intervention held an implicit theory of stability and participants in the effective intervention held an implicit theory of change—in both cases, the participants were correct and their retrospective accounts were as accurate or superior to their prospective accounts. Sprangers and Hoogstraten further demonstrated a pretest sensitization effect in their first experiment (using a Solomon four-group design) and recommended possibly discarding self-reported pretests in favor of self-reported thentests.

Howard, Ralph, et al. (1979) conducted a fifth study to determine if differences between pretest and thentest scores on a questionnaire about students' perceptions of their helping skills (HQ) could be the result of systematic memory distortion, instead of the hypothesized, conscious response shift. Undergraduate participants (*N=51*) enrolled in a helping skills course were randomly assigned to one of three groups: pre-post, post&then, and pre-post&then. Only students who completed all portions of the study (*n=37*) were included in analyses. After the pre-post and pre-post&then groups took the pre-survey, all students were randomly paired to interview each other and utilize helping skills (recorded helper-helpee interviews). At the end of the semester, after finishing the post&then survey, students in the pre-post&then group completed a memory HQ that asked them to remember and report responses they had chosen on the pre-survey. After all surveys were completed, students were randomly paired and participated in recorded helper-helpee interviews.

Study-blind raters rated each helper on three scales (feeling, content, and global). As suspected, mean memory ratings on the HQ were significantly different from then-survey ratings, but not significantly different from pre-survey ratings, signifying the response shifts were due to something other than systematic memory distortion. Significant differences between pretest and posttest behavioral measures were more closely aligned with significant differences between post/then subjective measures than with pre-post subjective measures, and subjective thentest measures were consistently lower than subjective pretest measures. In other words, like the previous studies, the significant treatment effect was better captured with a post/then design than a pre-post design.

Impression Management Theory

According to impression management theory, people provide consciously or subconsciously biased retrospective self-assessments in an attempt to make others perceive them more favorably (Hill & Betz, 2005; Nimon, 2014). For example, study participants may wish to show that they have always had the qualities that are being assessed (a "consistency effect" leading to matching posttest and thentest scores) or they may wish to show they have improved (exhibiting response shift through an over- or underestimation of thentest scores) (M. Ross, 1989). The fact that this phenomenon has been shown with control and intervention participants adds merit to the idea that humans have a natural tendency to present themselves in the most positive way possible (Hill & Betz, 2005).

In their evaluation of an intervention meant to improve parenting behaviors (*N=177*), Hill and Betz (2005) demonstrated that, while both pretests and thentests have biases, thentest ratings may increase impression management biases that eventuate in inflated posttest minus thentest change scores. The 100 study participants that completed all pre-post&then measures about intervention-related parenting behaviors and parent-child relationships were included in the analyses. Posttest minus pretest change scores and posttest minus thentest change scores showed program effectiveness, but posttest minus thentest change scores were significantly higher than posttest minus pretest change scores (i.e., response shift). Because pre-post effect sizes were similar to those in studies based on the same intervention and then-post effect sizes were much higher, the researchers claimed thentests might exaggerate the impact of an intervention.

To test for impression management biases in thentests, Hill and Betz (2005) had modified the survey instruments so half of the items were phrased as positive parenting behaviors or relationships and half were phrased negatively. On the thentest items, but not the pretest or posttest items, there was a significant difference between socially desirable and socially undesirable items. Average thentest behavior ratings were significantly lower than average pretest behavior ratings and average thentest relationship ratings were the same as average pretest relationship ratings. The researchers concluded that items phrased as undesirable and items about intervention-targeted behaviors are more prone to response shift due to impression management when reported on a thentest than items phrased as desirable and general attitudinal items.

Because Hill and Betz's (2005) study results ran contrary to findings by other researchers—that attitudinal items are generally more prone to response shift than behavioral items—they maintained that response shift in their study was due to cognitive distortion, not response shift bias. After discussing the pros and cons of pretests and thentests, the authors recommended researchers use prospective measures to determine mean program effects when the variables of interest are socially desirable behaviors and/or intervention-targeted behaviors and retrospective measures to determine subjective program effects on participants. They also emphasized the importance of including a thorough report of biases inherent in the researcher's chosen method.

Howard, Millham, Slaten, and O'Donnell (1981) conducted an assertiveness training study with college students *(N=40)* to investigate the claim that retrospective pretests introduce more impression management biases than traditional pretests. The researchers collected pretest and posttest data with four self-report measures: an assertiveness measure (CSES), a learning skills measure (LSQ), a social desirability measure (J-K), and a personal goals measure (COI). For the COI, each participant specified six traits he or she wished to develop (e.g., "assertion"), operationalized each trait with a behavioral definition (e.g., "initiating conversations with coworkers before work"), ranked the chosen traits in order of personal importance, and provided a pretest rating of his or her functioning on each trait. The participants' goals were woven throughout the training. After the treatment, without being told each participant's COI pretest ratings, the training facilitator estimated how much each participant gained on the participant's personal goals from the training. The researchers also collected thentest data for the CSES, LSQ, and COI. Eighteen students in the control group and 18 students in the training group completed the study and were included in the analyses.

Change scores using thentest ratings showed stronger treatment effects and were more highly correlated with the objective measures (the facilitator's COI ratings) than change scores using pretest ratings, signifying that the retrospective pretest may be more valid than the traditional pretest for the self-report measures used in their study. During post-intervention data collection, half of the control participants and half of the treatment participants were assigned to a "bogus pipeline"[28] group. Through an elaborate deception, the bogus pipeline participants were led to believe a voice analyzer was able to distinguish between their true and false verbal responses on the posttest and thentest questions, thereby lessening the likelihood of bogus pipeline participants providing socially desirable falsehoods.

On the intervention-targeted surveys (CSES and COI), there was evidence of socially desirable reporting of assertiveness (low to moderate relationship) on the pre-surveys, but not on the then-surveys; thentest scores of assertiveness exhibited less social desirability bias. The researchers did not find response shift on the nontreatment-related self-report measure (LSQ) and this supports the claim that respondents were reporting legitimate then ratings as opposed to ratings shifted by impression management biases. The researchers did not find an effect of the bogus pipeline deception on response shifts

---

[28] Howard et al. (1981) verified the utility of this particular bogus pipeline deception in a pilot study. Bogus pipeline participants from the pilot and larger study indicated, in a post-study debriefing, that they had been successfully fooled by the deception.

for intervention participants; meaning impression management was not responsible for response shifts. In sum, the treatment improved assertiveness and Howard et al. (1981) found that the retrospective prestests were simultaneously less vulnerable to social desirability bias and more accurate than the traditional pretests in this educational training. The authors recommended adding thentests to pre-post designs to account for response shift bias effects and obtain a more sensitive assessment of change.

Retrospective Pretest Designs

Once a researcher decides to include a retrospective pretest, he or she must decide whether or not to include a pretest and determine whether to administer the thentest on same questionnaire as the posttest (post&then) or on a questionnaire separate from the posttest (post-then). To clarify the distinction between thentest locations, consider a Likert scale survey with questions that ask for the respondent's level of agreement with various survey items on a scale ranging from strongly disagree to strongly agree. Commonly, post&then questionnaires use three separate columns to display survey items, posttest response options, and thentest response options. The survey items are in the leftmost column, the response options for how the participant views him or herself now (post) are in the middle column, and the response options for how the participant viewed him or herself prior to the intervention (then) are in the rightmost column. On some post&then questionnaires, survey items are in the middle column and the "now" response options are in the far left column. For a post-then administration, the respondents would first be presented with the survey items and asked to rate their level of agreement with the statements "now". Following this, the respondents would be presented with the same survey items on a separate form and asked to rate how much they agreed with the statements "then" (i.e., pre-intervention). Regardless of whether one uses a post&then or post-then layout, the majority of research supports laying out posttest items prior to thentest items (Klatt & Taylor-Powell, 2005).

Howard, Ralph, et al. (1979) recommended using a post&then procedure where respondents first report their present state (post) and then answer the same item "in reference to how they now perceived themselves to have been just before the [intervention] was conducted (Then)….in relation to the corresponding Post response to insure that both responses would be made from the same prospective" (p. 5). This post&then administration is problematic because it overtly signals an expected change and allows participants to reconstruct their current state in light of their perceived change (Nimon, 2014; Nimon et al., 2011a), and soliciting beliefs about perceived change can produce impression management and implicit theory biases (Lam & Bengo, 2003).

Nimon, Zigarmi, and Allen (2011b) compared the four types of retrospective pretest designs in an leadership training program *(n=139).* They used a self-report measure of perceived managerial competence and an objective measure of leadership performance to determine if the levels of criterion validity of thentest ratings and if measures of program effectiveness differ by type of research design. Validity estimates for thentest ratings were more comparable to validity data from a previous study when using the post-then format than when using the pre&then format. They also found that using separate posttest and thentest questionnaires provided a less biased measure of program effectiveness than using a single questionnaire (i.e., post&then designs exaggerated program effects). Slight pretest effects varied by post-then and post&then survey formats—there was a positive self-report pretest effect on data from the former and a negative self-report pretest effect on data from the latter. The researchers tentatively suggested the effect on pre-post-then data and pre-post&then data could be linked to a sensitization effect or compliance with implicit demands, respectively. The authors recommended separate administration of posttest and thentest surveys, and inhibiting participants from referring back to the posttest when completing the thentest. Based on their findings and recommendations by Mezoff (1981) and Sprangers (1996), they also suggested instructing participants to not worry about remembering whether or not their assessment of their pre-intervention state on the thentest matched their pretest ratings.

Summary

All self-report measures are prone to biases, including many that are beyond the scope of this review of the literature (e.g., introspection, reference). Germane to this dissertation study, varying the type instrument used to assess a subject's pre-intervention state (pretest or thentest) may vary the types of biases elicited. Validity concerns for thentest ratings may be subject to memory distortion, such as personal recall or adherence to an implicit theory. Thentests may also be vulnerable to biased reports when a subject wishes to appear favorably (better or the same as his or her past self). Response shift bias threatens the internal validity of self-report prospective measures and the potential for this threat "is exacerbated when a purpose of the treatment is to change the subjects' understanding or awareness of the variable being measured" (Howard, Ralph, et al., 1979, p. 1). Results from studies with prospective measures may also be muddied because a pretest may signal the goals of an intervention, influencing posttest and thentest scores (Sprangers & Hoogstraten, 1989). The main concern for using prospective measures is the underestimation of treatment effects and the main concern for using retrospective measures is the overestimation of treatment effects (Hill & Betz, 2005).

Retrospective pretests have shown much promise in measuring change while attenuating the effect of response shift bias. Post-then designs can economically provide a wealth of information when pre-measures are unavailable or undesirable (Hill & Betz, 2005). Pre-post-then designs may help researchers uncover true and/or perceived change. A researcher's choice to select one method over another should be based on the goals of the research, determination of which biases are more likely, and a decision about which biases are most crucial to avoid. As demonstrated in the studies above, depending on which theory the researcher ascribes to, different conclusions can be drawn about the same data—evidence of response shift may validate the response shift bias theorist's claim that thentests are more valid and concurrently validate the implicit theorist's or impression management theorist's claim that prospective assessments are more accurate (Nimon, 2007; Norman, 2003). Many researchers suggest including an independent measure of change to help determine which theory supports conclusions drawn (Allen & Nimon, 2007; Hill & Betz, 2005; Sprangers & Hoogstraten, 1989). When concurrent validity measures, such as objective or qualitative data, are unavailable and/or rigorous study designs are not feasible, it is important for researchers to consider all possible interpretations of study results.

Document Roadmap

In Chapter 2, I defined dm-noncognitive factors, and I provided supporting research for the dm-noncognitive factors used in my dissertation research: math equanimity, math and college belongingness, math self-efficacy, and math mindset. I described math concept transfer and the difficulties associated with uncovering whether or not students are prepared to learn and innovatively utilize their knowledge. I detailed ways in which the Foundations curriculum supports the development of dm-noncognitive factors and promotes productive persistence and transfer. I ended the chapter with research about methodological issues concerning survey data, while highlighting the threat of response shift bias and outlining the potential benefits of retrospective pretests. In Chapter 3, I discuss my research methods and plans for data analysis.

Chapter 3: Methodology

This is an exploratory mixed methods study that focuses on practical ways to discern evidence of select dm-noncognitive factors of developmental mathematics college students enrolled in The NMP's Foundations of Mathematical Reasoning course, and how those factors relate to the students' course success and ability to transfer knowledge to proximal mathematical tasks. I used self-report surveys to look for changes in the students' math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness; and compared the survey results with semester outcomes. The semester outcomes include math course grade, final exam grade, and math course percent attendance. I also collected student self-reported demographic data to control for student background characteristics. I examined the usefulness of short, retrospective pre-surveys for measuring dm-noncognitive factors of developmental mathematics students.

A subset of students took the online Developmental Assessment of Place Value Understanding (DAPVU)—a modified version of the Assessment of Place Value Understanding (APVU)—as a quantitative semester outcome measure and as a qualitative transfer measure. I did not receive sufficient data on the DAPVU to use it as a quantitative outcome measure. I requested interviews with students to provide a more complete picture of the impact of the course on students' dm-noncognitive factors and ability to transfer place value concepts, but only a few students agreed. Results from the DAPVU and interviews function as supporting qualitative information.

This chapter includes my research questions and hypotheses, and descriptions of my study participants, data collection measures, data collection procedures, and data analysis methods.

RESEARCH QUESTIONS AND HYPOTHESES

In the following research questions, the dm-noncognitive factors of interest are math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness. The semester outcomes[29] of interest are math course grade, math final exam score, and math course attendance.

1. Do students exhibit differences over time in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness?

   RQ1A. Do students exhibit beginning-of-semester to end-of-semester differences in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness?

   RQ1A Research Hypothesis: Students will exhibit beginning-of-semester to end-of-semester improvements in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness.

   RQ1B. Do students exhibit pre-survey to then-survey differences in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness?

   RQ1B Research Hypothesis: Students will exhibit pre-survey to then-survey differences in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness.

2. Do beginning-of-semester to end-of-semester differences in students' math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness predict semester outcomes (math course grade, math final exam grade,

[29] I was unable to use Developmental Assessment of Place Value Understanding scores as a quantitative outcome measure.

math course percent attendance, and Developmental Assessment of Place Value Understanding score)?

RQ2 Research Hypothesis: Beginning-of-semester to end-of-semester differences in students' math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness will predict semester outcomes.

3. Do students exhibit evidence of their ability to transfer their knowledge to novel place value problems?

   RQ3 Research Hypothesis: Students will exhibit evidence of their ability to transfer their knowledge to novel place value problems.

PARTICIPANTS

The participants were community college students enrolled in the Foundations of Mathematical Reasoning course during the Fall 2015 semester that was created as part of the New Mathways Project. I chose study participants from two mid-sized southern community colleges (College A and College B) that had moved to full-scale implementation of the course. I obtained Institutional Review Board approval to collect data at College A and College B (See Appendix A). I contacted administrators at these two colleges to gain access to instructors and based my convenience sample on instructor willingness to assist with the study. College A offered an internet-only version and an 8-week version of the course in addition to offering the course for the full semester on campus. I only recruited students from full semester courses that met in person. College A offered 19 such sections taught by 12 instructors with approximately 20 students each and College B offered 14 sections taught by 9 instructors with approximately 25 students each. All instructors for these 33 sections agreed to assist with the study. I included all students from these sections who gave their consent to participate. The consent description is in Appendix B and the consent form is in Appendix C. A total of 597 students participated in this study[30], with 292 from College A and 305 from College B. The following demographic information about my sample was reported by study participants on a survey given at the end of the semester. I listed the number of respondents that provided information for each question in parentheses. Distributions of demographic variables are also listed in tabular form in Appendix I.

Approximately 70% reported as female and 30% male (*n*=467). For enrollment status, 40% were considered part time students and 60% were full time students (*n*=329). Seventy-one percent reported having no dependents, 10% reported having 1 dependent, 10% reported having 2 dependents, and 9% reported having 3 or more dependents (*n*=325). Participants were predominately native English speakers, with only 9% reporting that English was not their native language (*n*=319). Half of the students reported that they or one of their siblings was a first generation college student (*n*=319). Participants described themselves as Hispanic, Latino, or Spanish (37.1%); White, Non-Hispanic (35.6%); Black or African American, Non-Hispanic (16.2%); Asian, Asian American, or Pacific Islander (2.7%); American Indian or other Native American (1.2%); Native Hawaiian (0.6%); or Other/Multi-racial (6.6%) (*n*=334).

For the question about maternal education, approximately 12% reported that their mother did not graduate from high school, 33% reported their mother completed high school or received a GED, 23% reported that their mother attended, but did not graduate from, college, 13% reported that their mother received an associate degree, 8% reported that their mother received a bachelor's degree, 4% reported that their mother received a master's degree, 1% reported that their mother received a doctoral degree, and 6%

[30] This is a consent rate of approximately 80%.

reported that their mother's education was not reflected in the options given or they did not know their mother's level of education (*n*=325).

Approximately 55% of the respondents were between 18 and 19 years of age; the other categories were: under 18 years (1.7%), 20 to 21 years (9.6%), 22 to 24 years (6.8%), 25 to 29 years (8.5%), 30 to 39 years (10.7%), 40 to 49 years (4.9%), 50 to 64 years (0.2%) (*n*=468). The amount of hours students worked for pay each week were distributed with 35.7% not working, 14.4% working less than 20 hours, 15% working from 20 to 25 hours, 4.8% working from 26 to 30 hours, 8.7% working from 31 to 35 hours, 5.4% working from 36 to 39 hours, and 15.9% working 40 hours or more per week (*n*=333).

DATA COLLECTION MEASURES

Key Variables

The key noncognitive factor variables are: math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness. These items are measured on student self-report pre-surveys and post-then-surveys (described below), and treated as continuous variables because there are five or more response options for each question.

The semester outcome variables are: math course grade, math final exam grade, and math course percent attendance. The instructors provided course grades and final exam grades on a scale of 0 to 100. The instructors also provided the number of days each student attended and the total number of class days. I used this information to calculate a percentage of attendance for each student, so each student has an attendance score on a scale of 0 to 100. I was originally going to use the DAPVU as an outcome measure. Below, I describe the DAPVU and why I was ultimately unable to use it as an outcome measure.

The demographic variables are: gender, age group, race/ethnicity, enrollment status, job status, number of dependents, English as native language, maternal education, and first generation college student.

Surveys

Participants took a pre-survey about dm-noncognitive factors at the beginning of the semester and a related post-then-survey before the last drop date near the end of the semester. At the end of the post-then-survey participants were asked to complete a small set of demographic questions, as demographic backgrounds may be related to the results. The pre-survey is in Appendix E and the post-then-survey is in Appendix G. Pre-, post-, and then-survey response rates for dm-noncognitive factors are in Appendix H. Demographic survey response rates are in Appendix I.

Survey Design

Well-designed surveys provide researchers the opportunity to collect and analyze data quickly and efficiently, but, as is the case with other data collection methods, they have inherent biases. When possible, I attempted to diminish biases and response effects in the surveys used in this study. I aimed to create practical measures that followed recommendations of researchers familiar with survey design among low-education respondents, and I have discussed some of these considerations. In addition to being statistically valid and reliable, surveys given to low education respondents need to be face valid, brief, and clear (Yeager, Bryk, et al., 2013). Survey brevity also reduces the level of class disruption and decreases the chance of survey fatigue. The pre-survey used in this study has fourteen questions, the post-then-survey has twenty-eight questions, and the demographic survey has nine questions.

Participants were offered incentives and the surveys were administered during class time to reduce the chance of nonresponse and voluntary response bias (Tourangeau, 2004). Because I had a convenience sample, I was unable to account for undercoverage bias. To deter social desirability bias, participants were reminded of the confidentiality of their responses and were told that there were no right or wrong answers. For the ordinal, dm-noncognitive factor items, I used a linear layout of response options so respondents could more easily identify where they best fit on each continuum (Dillman & Christian, 2002). I labeled all of the scale points with words, not just the end points, to clarify what each point signifies and improve validity and reliability (Dillman & Christian, 2002; Krosnick, 1999; Krosnick & Presser, 2010). This was especially important because scale points were labeled with letters instead of numbers, due to the use of scantrons.

Research has shown that surveys with agree/disagree rating scales are more prone to acquiescence response bias, where respondents are more inclined to agree than disagree regardless of the item's content, than surveys with item specific or construct specific response options[31] (Saris, Revilla, Krosnick, & Shaeffer, 2010; Yeager, Bryk, et al., 2013). Further, acquiescence response bias may be strongest among low education respondents (Abts, 2012). In order to mitigate acquiescence response bias, I looked for scales that had construct specific response options. The math equanimity scale and math belongingness scale meet this requirement. The math theory of intelligence scale uses the

[31] Yeager, Bryk, et al. (2013) provide example formats to delineate the difference: The survey item could be posed "in agree/disagree format as 'I would feel anxious taking a math or statistics test' (response options: 1=*strongly disagree*, 5=*strongly agree*), or it could be written in a construct-specific format as, 'How anxious would you feel taking a math or statistics test?' (response options: 1=*not at all anxious*, 5=*extremely anxious*)" (p. 25).

agree/disagree format, but Yeager, Bryk, et al. (2013) found it to have the same or better predictive validity compared to an item specific format. I was unable to find a construct specific self-efficacy scale, despite an extensive database search (e.g., Google Scholar, EBSCO, etc.). I chose a self-efficacy scale, from the Patterns of Adaptive Learning Survey, that had consistently been proven to be valid and reliable with diverse populations (Linnenbrink & Pintrich, 2002; Midgley et al., 2000; M. E. Ross, Shannon, Salisbury-Glennon, & Guarino, 2002).

One purpose of this study was to determine the practicality of a retrospective pretest to account for response shift bias (Howard, 1980; Howard, Ralph, et al., 1979). I used a pre-post-then-survey design (Allen & Nimon, 2007; Coulter, 2012; Howard, Dailey, et al., 1979; Nimon, 2007; Nimon et al., 2011b). The post-then-survey has three sections in this order: post, then, and demographic questions. Students were instructed not to begin the then-survey until they had completed the post-survey and they were instructed not to begin the demographic section until they had completed the then-survey. The directions for the pre and post-then-surveys about dm-noncognitive factors are: “For each question, please completely bubble in only one response. There are no right or wrong responses, only the way you feel about each statement or question.” Additional instructions for the then-survey asked students to think back to the beginning of the semester and choose responses that would have best described them at the beginning of the semester. The wording was based on guidelines suggested by Mezoff (1981) and adapted from a retrospective pre-survey that was used in a study that compared four survey formats: pre-post-then, pre-post&then, post-then, and post&then (Nimon, 2007; Nimon et al., 2011b).

DM-Noncognitive Factors

The pre-post-then-surveys were designed to measure five latent dm-noncognitive factor constructs: math equanimity, math mindset, and math self-efficacy, math belongingness, and college belongingness. Higher scores represent stronger dm-noncognitive factors.

The *math equanimity* scale and the *belongingness* measures came from a practical measure that was designed specifically for students in developmental math courses (Yeager, Bryk, et al., 2013). In line with the goals of creating a practical measure, the authors created an instrument that was brief, highly predictive, face valid (to faculty and students), and sensitive to changes in the short term. The authors provide a thorough description of how they validated their instrument and discuss the inappropriateness of using typical psychometric summary statistics to evaluate practical measures. For example, they explain how it would be incongruous to assess the internal consistency reliability of a practical measure because such measures are designed to provide strong predictive analytics with very few items, most of which are intentionally non-overlapping and minimally correlated. To provide support for their claims, the authors reference Cronbach (1961) as saying, “If predictive validity is satisfactory, low reliability does not discourage us from using the test” (p. 128). The key point is—practical measures are designed to assess and improve educational experiences, not for theory development, and, while they should be held to equivalently high standards as measures that are intended for theory development, they should be validated in an entirely different manner.

The practical measure’s math equanimity scale consists of four questions that ask how anxious you would feel “listening to a lecture in math class,” “taking a math test,” “signing up for a course in math,” and “the moment before you got a math test back.”

The response options range from (a) extremely anxious to (e) not at all anxious. So, a high score represents math equanimity and a low score represents math anxiety. I took the unweighted average of the completed items to generate a composite score for each student. In other words, if a student only responds to two of the four items, these items are averaged to calculate the students' math equanimity score. There are two belongingness measures—college belongingness and math belongingness—each on a scale from (a) always to (e) never: "When you think about your [college or math class], how often, if ever, do you wonder: 'Maybe I don't belong here?'" These items were analyzed separately. On Yeager and Bryk's (Yeager, Bryk, et al., 2013) practical measure, only the college item was used and it was shown to be a very strong predictor of course completion and course passing. This outcome was replicated in large samples across colleges. When administered in the fourth week of class, this item has been shown to be the single best predictor of course outcome, even after controlling for demographics and prior math knowledge and (Yeager, Bryk, et al., 2013).

I assessed students' math mindsets using the *math implicit theory of intelligence* scale. The three-item implicit theory scale has a high internal consistency ($\alpha \geq .94$) and test-retest reliability at two weeks ($\alpha \geq .80$) (Dweck, Chiu, & Hong, 1995). Students were asked how much they agree or disagree, on a scale from (a) strongly agree to (f) strongly disagree, for each of the following: "You have a certain amount of math intelligence and you really can't do much to change it," "Your math intelligence is something about you that you can't change very much," and "You can learn new things, but you can't really change your basic math intelligence." Each respondents' score for the math implicit theory of intelligence scale was calculated in accordance with Dweck, Chiu, and Hong's (1995) recommendation. I averaged the completed implicit theory of intelligence items to assign an overall implicit theory score for each student. Persons with a score of 3.0 or below are said to be holding an entity theory of math intelligence (fixed mindset) and those with a score of 4.0 or above are holding an incremental view (growth mindset). Persons in the range between 3.0 and 4.0 are generally not labeled as either entity or incremental theorists.

The *math self-efficacy* scale was taken from the Patterns of Adaptive Learning Survey (PALS) (Midgley et al., 2000). This is a well-established instrument that has been validated at many different age levels, ranging from elementary to college-age students (Linnenbrink & Pintrich, 2002; Midgley et al., 2000). According to Midgley et al. (2000), the internal consistency reliability (coefficient alpha) for this construct was 0.78 when used with younger elementary-aged children. The authors reported the reliability coefficients for younger children whenever possible because the coefficients with older children are usually higher. Linnenbrink and Pintrich (2002) recommend adjusting motivation items such as these to make them context specific; students' responses on a self-efficacy scale in one domain may differ from their responses on a self-efficacy scale in a different domain. It has been shown that changing the wording of these self-efficacy items to make them domain specific does not decrease the reliability of this scale, so I changed the wording of the survey items to reflect math self-efficacy (Meuschke, 2005). The five items that make up the scale are as follows: "I'm certain I can master the skills taught in my math class this semester," "I'm certain I can figure out how to do the most difficult work in my math class," "I can do almost all the work in my math class if I don't give up," "Even if the work in my math class is hard, I can learn it," and "I can do even the hardest work in my math class if I try." Students indicated how much each statement was true of them with responses ranging from (a) not at all true of me to (e) very true of me. I created composite scores by taking an unweighted average of the completed math self-efficacy items.

Demographics

I collected demographic data to describe my sample and control for demographic differences. The demographic section of the survey follows immediately after the “then” section of the post-then-survey. They are placed at the end of the survey to reduce the chance of stereotype threat (Steele, 1997; Steele & Aronson, 1995). It includes nine questions and response options that are commonly used on national surveys of community college students and surveys specific to developmental math students.

Students were asked to report their gender (male or female), age group (measured in years), race/ethnicity (American Indian or other Native American; Asian, Asian American, or Pacific Islander; Native Hawaiian; Black or African American, Non-Hispanic; White, Non-Hispanic; Hispanic, Latino, or Spanish; Other/Multi-racial, enrollment status (part time or full-time), job status (measured in number of hours worked per week), number of dependents (0, 1, 2, or 3+), and whether or not English is their native language (yes or no). There are two questions related to their family’s educational background: maternal education (high school graduate; high school diploma or GED; some college, did not complete degree; associate degree; bachelor’s degree; master’s degree/1st professional; doctoral degree; unknown or other) and whether or not they or one of their siblings is a first generation college student (yes or no).

Developmental Assessment of Place Value Understanding

At the end of the semester, all consented participants that completed the pre-, post-, and then-surveys were invited to take the Developmental Assessment of Place

Value Understanding (DAPVU) online to measure their ability to transfer their place value knowledge to novel situations. The DAPVU is an adaptation of the Assessment of Place Value Understanding (APVU); the thirteen-item[32] APVU was created and validated by Hannigan (1998) and Rusch (1997) for use in their dissertation research about preservice teachers' explicit place value understanding.

Rusch (1997) defines explicit understanding as a "precise understanding of the concept being addressed which can be clearly articulated and convincingly justified. Explicit understanding includes a clear understanding of the relationships among ideas, as well as the logical development of related algorithmic procedures and generalizations" (p. 97). Hannigan (1998) and Rusch (1997) composed a set of twelve crucial elements that characterize explicit place value understanding; I list these key elements in Appendix L. Hannigan makes the distinction that, although the APVU was used with preservice teachers, it was not an assessment of pedagogical content knowledge. It is appropriate to use a measure of explicit place value understanding as a measure of concept transfer with Foundations students because: place value is a fundamental mathematical concept, Foundations students have previously been required to solve numerous base-ten place value problems, and place value is not the focus of the Foundations curriculum. Hannigan explains, "understanding [as used in her research with Rusch] is not simply the amassing of knowledge but the ability to apply that knowledge to novel or unfamiliar situations," and a person equipped with such understanding "is flexible in their ability to work within and between representations" (p. 5). Hannigan also links her work with Rusch to Hatano's (1988) work, noting that "a well-connected knowledge structure" may help a

[32] The pre- and post-versions of the APVU were mostly parallel and each contained 13 items. However, there were two major differences between some problems on the pre- and posttest. Because of these differences, Hannigan claimed there were fourteen unique tasks across the pre/posttest.

person utilize his or her adaptive expertise when solving problems within or related to the domain of place value (p. 24). The APVU is more than a measure of replicative or applicative knowing, but also a measure of interpretive knowing.

Place value is a "positional notation scheme for quantity" (Hannigan, 1998, p. 36) and "the grouping scheme used in a representation of quantity is one of the, if not *the*, most essential element" (p. 38). Rusch (1997) and Hannigan found that place value problems involve contexts that are either familiar or unfamiliar and bases (grouping schemes) that are either systematic or nonsystematic. Hannigan refers to the four possible combinations of these bases and contexts (i.e., familiar-systematic, familiar-nonsystematic, unfamiliar-systematic, and unfamiliar-nonsystematic) as quantitative representations and I use this same terminology. Table 1 explains the four quantitative representations further.

Table 1

Classification of Quantitative Representations

	Systematic Systematic grouping schemes are those in which the number of items required to create one group of the next place value remains constant.	**Nonsystematic** Nonsystematic grouping schemes are those in which the number of items required to create one group of the next place value varies from place to place.
Familiar A familiar context is one that most developmental math students have come across in both traditional mathematics curriculum and in daily living routines.	The Base-10 System Metric Measurement	Time: Seconds, Minutes, Hours, Days, Weeks, Years, … Imperial Measurement: Inches, Feet, Yards, … Cups, Pints, Quarts, gallons, … American Coinage: Penny, Nickel, Dime, Quarter, …
Unfamiliar An unfamiliar context is one which uses a place value structure, but which developmental math students may <u>not</u> have come across in either their school years or their daily life.	Base-n systems ($n \neq 10$): Base-2, Base-5, Base-16, …	Foreign Coinage: Former British coin system which included pence, shilling, pound, …

Note. This table is adapted from Rusch's (1997) Matrix of Place Value Grouping Schemes (p. 38) and Hannigan's (1998) Classification of Quantitative Representations (p. 46).

The authors of the APVU tried to create a balanced mix of problem situations, based on quantitative representations (e.g., familiar-nonsystematic) and operation or structure (e.g., addition problem situation, conversion problem situation). There were only two multiplication problem situations and no division problem situations because the authors believed explicit place value understanding could be adequately demonstrated using structure, addition, and subtraction problem situations and because meaningful multiplication and division problem situations were much more difficult to find or create.

The APVU has rubrics for seven knowledge dimensions: Error Reproduction (ER), Depth of Analysis (DA), Use of Descriptive Language (DL), Choice of Correct Representation (CR), Accurate Computation (AC), Analysis of Computation Method (CM), and Analysis of Symbolic Representation (SR). There are at least two rubrics associated with each of the thirteen problem situations. There are five levels for each knowledge dimension rubric. Level one and two scores represent algorithmic understanding, level three scores represent tacit understanding, level four is explicit understanding, and level five is representative of clearly explicit understanding. Rusch (1997) describes algorithmic, tacit, and explicit understanding as follows:

> Algorithmic: Knowledge of how to manipulate symbols to get an "answer." This level of mathematical knowledge does not necessarily imply understanding of why the algorithmic steps make sense.
>
> Tacit: Algorithmic knowledge with some intuitive understanding of the logical foundation from which the concept or algorithm emerges. This understanding is, however, somewhat vague and as a result it is difficult for the individual to articulate the logic which brings meaning to the concept.
>
> Explicit: Precise understanding of the concept being addressed which can be clearly articulated and convincingly justified. Explict understanding includes a clear understanding of the connections between the concept being addressed and

> related concepts, as well as an ability to articulate the logical development of related algorithmic procedures and generalizations. (p. 34)

I will provide more details about how Hannigan (1998) and Rusch (1997) analyzed problem situations with the knowledge dimension rubrics in the section about scoring the DAPVU.

Constructing the DAPVU

While the authors of the APVU attempted to include a balanced mix of problem situations, unfamiliar-systematic problem situations were overrepresented on the APVU. Three of the unfamiliar-systematic problem situations on the APVU involve explicit use of the standard algorithm for computation in different bases. A standard algorithm, by definition, is something that must be conventionally taught. Because Foundations students may have never been introduced to the standard algorithm, these problem situations are not appropriate for measuring Foundations students' ability to transfer place value concepts and were not included in the DAPVU. Hannigan (1998) and Rusch (1997) mentioned several concerns about the notational structure, unfamiliar-systematic problem situation—it was confusing to students, it didn't add substantial information to the research, and it needs extensive revisions in order to be useful. This problem situation was excluded from the DAPVU.

I changed some of the names used in the APVU problem situations so students may be less likely to find the original assessment online (e.g., changed "Picabo's Race Times" to "Pat's Skiing Competition"). I made minor modifications to the framing of some problem situations so as to avoid confusion related to uncommon words and items. For example, the APVU references “runs” when describing race times in a skiing competition and describes a teacher putting math “chips” on a white board. The DAPVU, instead, refers to ski times down a hill (making the problem more accessible to people of low socioeconomic status) and a teacher drawing squares on the board (making the problem more accessible to people who are unfamiliar with mathematics manipulatives). I believe the problem situations were substantively the same after these changes.

Hannigan (1998) and Rusch (1997) provided recommendations for revisions in their dissertations and I used some of these recommendations in my adaptation of the APVU. According to Hannigan, some wording needed to be changed in some of the problems “to help the students focus on analyzing conceptual understanding rather than procedural understanding” (p. 104). Following this recommendation, on the Maria Error Pattern problem situation, I changed “briefly describe your assessment of what [Maria] does not understand” to “describe your assessment of what concept or concepts Maria does not understand. Try to explain what she does not understand, not just what procedure she is using.” On APVU problems that instruct, “show all your work,” the DAPVU instructs: “Describe, in detail, exactly what you did to solve the problem. Please be very specific because I am not able to see your work.” This addition was two-fold: it could help me determine what I could not see of their scratch-work (evidence of replicative and applicative knowing) and it could provide insight into their thought process (evidence of their interpretive knowing). In the hope of further surfacing student understanding and to make the DAPVU closer to a PFL measure that considers what students “know with”, I added prompts such as: “If you are not sure how to find the combined time, what ideas do you have about the problem? Are there any questions you would like answered to help you solve the problem?”

The authors of the APVU designed it so the average student could complete it in less than one hour. The first version of the DAPVU included all APVU problem situations, except the three standard algorithm problem situations and the notational structure, unfamiliar-systematic problem situation. As mentioned above, I included prompts to determine what procedures students used in their scratch-work and better surface students’ preparation for future learning. I asked friends and family members of various ages and educational backgrounds to take the first version of the online DAPVU to check for clarity and time consumption. The first version took respondents anywhere from 1.5 to 4.5 hours to complete. Most people wrote something for all problem situations, even if it was just explaining that they did not know how to approach the problem situation. Because these are people I know personally, it is likely that they spent more time on the assessment to provide the best feedback possible. Still, the fact that 1.5 hours was the minimum amount of time spent on the assessment indicated that it was excessively long. One mathematician that took the assessment commented that it was a “place value workout” and I would be “making them work really hard for that gift card.” I needed to find a balance where it would be difficult enough to give people an opportunity productively persist, but not so time consuming that they would merely be frustrated by the process and quit.

I consulted with three mathematicians and a mathematics education doctoral student to choose a representative subset of problem situations to reduce the length of the assessment for the final version of the online DAPVU and I ultimately decided to keep five problem situations (more details below). I also consulted with the experts on framing the problem situations so the prompts could more accurately assess students' interpretive knowing. The four experts agreed that the revised version of the online DAPVU could adequately differentiate where students lie on the spectrum from tacit to clearly explicit place value understanding, as defined by Hannigan (1998) and Rusch (1997), and provide some insight into the students' interpretive knowing. While the experts agreed the five-problem version of the DAPVU was parallel to the APVU, two of the mathematicians were apprehensive about the usefulness of the notational, familiar-systematic problem situation (Bobby's Squares). One said it may be "too vague" and another "caution[ed] against reading too much into" a person's response because not doing well on the problem situation may be indicative of something other than lack of place value understanding. I decided to keep the Bobby problem situation because it was the only notational problem, it did not add much length to the assessment, and several friends and family members talked about place value in their responses to the problem situation. There is a chance it may not be appropriate to include the Bobby problem situation in a person's overall quantitative DAPVU score, but it may provide some qualitative insight.

I asked more friends and family members of various ages and educational backgrounds to take the five-problem, online version of the DAPVU to check for clarity and timing. The average person took approximately one hour. Based on feedback, I made minor wording changes to the prompts where I ask them to explain their work and the respondents agreed that the changes added clarity. The mathematicians and mathematics education doctoral student agreed that the online DAPVU was ready to be used for my research purposes. The DAPVU is in Appendix M[33].

In Table 2, I compare the mix of problem situations on the APVU and the shorter, final version of the DAPVU. The first number represents the number of APVU problem situations and the number in parentheses represents DAPVU problem situations. The DAPVU consists of five distinct problem situations: two familiar-systematic, one familiar-nonsystematic, one unfamiliar-systematic, and one unfamiliar-nonsystematic. Two problem situations involve addition, two problem situations involve subtraction, and one problem situation involves notational understanding.

[33] The online DAPVU is in color and has one problem situation per page.

Table 2

Comparisons of Quantitative Representations to Operation/Structure for APVU and DAPVU Problem Situations

	Quantitative Representations			
Operation/ Structure	Familiar-Systematic	Familiar-Nonsystematic	Unfamiliar-Systematic	Unfamiliar-Nonsystematic
Notational	1 (1)		1 (0)	
Conversion			1 (0)	
Addition	1 (0)	1 (1)	1 (0)	1 (1)
Subtraction	1 (1)	1 (0)	2 (1)	1 (0)
Multiplication			1 (0)	1 (0)

Note. The numbers without parentheses represent the number of APVU problem situations per operation/structure by quantitative representation. The numbers in parentheses represent the number of DAPVU problem situations per operation/structure by quantitative representation (e.g., There is one unfamiliar-systematic addition problem on the APVU and no unfamiliar-systematic addition problems on the DAPVU).

DAPVU Problem Situations

Problem Situation 1 (PS1): Pat's Skiing Competition

PS1 is a familiar-nonsystematic addition problem that asks students to compute the amount of time it takes a skier to ski down a hill twice, given two times in minutes, seconds, and milliseconds (2:53.67 and 2:50.54). Students are also asked to describe what they did to solve the problem. If they are unsure about how to solve the problem they are asked to provide their ideas about how to solve the problem.

Problem Situation 2 (PS2): Bobby's Squares

PS2 is a familiar-systematic notational problem that asks students to evaluate the thinking of a hypothetical student, Bobby. In the scenario, Bobby's teacher draws 26 squares on the board and Bobby is able to count the squares and write the number 26.

When his teacher points to the six Bobby had written and asks Bobby to circle the squares represented by the six, Bobby circles six squares. When his teacher points to the two Bobby wrote and asks Bobby to draw a box around the squares represented by the two, Bobby draws a box around two squares. The students are asked to evaluate Bobby's thinking and report what Bobby understands. They are also asked if there is anything Bobby doesn't understand about the problem and, if so, to report what Bobby does not understand.

Problem Situation 3 (PS3): Chocolate Factory

PS3 is an unfamiliar-systematic subtraction problem. PS3 may be the best test of near transfer because unfamiliar-systematic problem situations are "similar enough to base-ten to allow concepts to transfer but unfamiliar enough so that disequilibrium will occur…and new knowledge develop" (Hannigan, 1998, p. 46). Students are presented with a scenario inside a chocolate factory where workers use base-four notation to represent cases of chocolates. In the factory, four singles complete one package, four packages complete one box, four boxes complete one carton, and four cartons complete one case. The students are provided with an example in factory notation: "A 1231 is a partially filled case that has 1 carton, 2 boxes, 3 packages, and 1 single chocolate." They are asked to record, in factory notation, how many chocolates they would need to fill a partially full case, where the partially full case is a 3012. They are also asked to describe what they did to solve the problem. If they are unsure about how to solve the problem they are asked to provide their ideas about how to solve the problem.

Problem Situation 4 (PS4): Rugolian Rug Merchant

PS4 is an unfamiliar-nonsystematic addition problem. Students are provided information about the currency denominations used in the fictional kingdom of Rugolia.

In the Rugolian money system, yellow is the basic coin, the green coin is equivalent to four yellow coins, the red coin is equivalent to two green coins, and the blue coin is equivalent to three red coins. The students are asked to add two rug prices and to record the sum with the fewest number of coins possible. They are also asked to describe what they did to solve the problem. If they are unsure about how to solve the problem they are asked to provide their ideas about how to solve the problem.

Problem Situation 5 (PS5): Maria's Error Pattern

PS5 is a familiar-systematic subtraction problem. Students are presented with four examples of a hypothetical student's subtraction work using the standard algorithm. Maria, the hypothetical student, solved two of the problems correctly and two incorrectly. Students are told that Maria's errors indicate there is something she doesn't understand. Then they are asked to provide the answers Maria would give for three other problems. If students accurately reproduce Maria's error pattern, one of the problems should be solved correctly and two of the problems should be solved incorrectly. The students are asked to describe what they did to come up with Maria's likely solutions and to provide an assessment of what concept or concepts Maria does not understand.

Scoring the DAPVU

The DAPVU uses the same knowledge dimension rubrics as the APVU, except the Choice of Correct Representation rubric. This rubric was only used to assess the notational structure, unfamiliar-systematic problem situation that was excluded due to numerous concerns by the APVU authors, and they specified that inclusion of that problem situation, as it was originally written, did not add much information to the assessment. The rubrics for each of the five DAPVU problem situations are in Appendix N. I broke down each DAPVU problem situation by operation or structure, quantitative representation, and knowledge dimension in Table 3. I further broke down each quantitative representation and knowledge dimension by the specific number of subtasks (shown in in parentheses). For example, the Familiar-Nonsystematic quantitative representation has two subtasks: Accurate Computation on Pat's Skiing Competition and Analysis of Computation Method on Pat's Skiing Competition. Similarly, the Error Reproduction knowledge dimension has one subtask: Maria's Error Pattern. The DAPVU has a total of 12 subtasks.

Table 3

Operation or Structure, Quantitative Representation, and Knowledge Dimensions of DAPVU

DAPVU Problem Situations	Operation or Structure	Quantitative Representation (12)	Knowledge Dimensions (12)					
			ER (1)	DA (2)	DL (2)	AC (3)	CM (3)	SR (1)
Maria's Error Pattern [James' Error Pattern]	Subtraction	Familiar-Systematic (5)	X	X	X			
Bobby's Squares [Bobby's Understanding]	Notational			X	X			
Pat's Skiing Competition [Picabo's Race Times]	Addition	Familiar-Nonsystematic (2)				X	X	
Chocolate Factory [Caramel Factory]	Subtraction	Unfamiliar-Systematic (2)				X	X	
Rugolian Rug Merchant [Zorandrian Rug Merchant]	Addition	Unfamiliar-Nonsystematic (3)				X	X	X

Note. The associated knowledge dimensions are: Error Reproduction (ER), Depth of Analysis (DA), Use of Descriptive Language (DL), Accurate Computation (AC), Analysis of Computation Method (CM), and Analysis of Symbolic Representation (SR). The number of subtasks for each of the knowledge dimensions and quantitative representations are in parentheses. The corresponding APVU tasks are in square brackets.

I originally planned to score the DAPVU as an SPS measure of transfer and more qualitatively as a PFL measure of transfer. To score the DAPVU as an SPS measure, I would use the scoring process Hannigan (1998) and Rusch (1997) used with the APVU. First, each problem situation would be scored using the rubrics on a level from one to five for each of the problem situation's knowledge dimensions. Consider an example using the problem situation about Bobby's notational understanding: To score a student's response, a grader would compare the response to the Depth of Analysis rubric and assign a score from one (algorithmic place value understanding) to five (clearly explicit place value understanding), based on the student's depth of analysis on the Bobby problem situation. Then the grader would compare the student's response to the Use of Descriptive Language rubric and assign another score from one to five based on the student's use of descriptive language in his or her response to the Bobby problem situation. The grader would score each of the remaining four problem situations in relation to the subtasks aligned with each problem situation. Refer back to Table 3 for clarification.

Hannigan (1998) and Rusch (1997) created the AVPU rubrics so that total scores could be computed for an individual on a knowledge dimension, quantitative representation, or across both. Because the DAPVU is shorter than the APVU, there are insufficient data points to accurately assess a student's understanding for each individual knowledge dimension and each individual quantitative representation. There is, however, a balanced mix of twelve subtasks to create an overall DAPVU score. Simple means and standard deviations may be used to create an overall DAPVU score for each individual by averaging each student's score on the twelve subtasks. As a result, each student would receive an overall DAPVU score ranging from one (algorithmic understanding) to five (clearly explicit understanding). The DAPVU also incudes prompts similar to those in the eagle challenge (described in section on transfer in Chapter 2) to provide students with opportunities to exhibit understanding that remains hidden in traditional SPS transfer measures. These prompts can help researchers determine DAPVU scores and provide information for deeper qualitative analyses.

Scoring the Online DAPVU

For logistical reasons, the DAPVU used in this study was administered online. After I received the online responses, I compared the DAPVU knowledge dimension rubrics to student responses to determine what information was unavailable as a result of online administration. Several of the DAPVU rubrics require access to, or thorough descriptions of, student work. For example, the levels in the Analysis of Symbolic Representation rubric distinguish between algebraic, symbolic, and pictorial representations. If a student uses tally marks, but does not divulge this pictorial representation, the evaluator is unable to assign an appropriate score. I asked students to describe their work, but their descriptions were not thorough enough for me to score the online DAPVU in the same manner as I could score a paper version of the DAPVU. As a result, all of my analyses of the online DAPVU are descriptive.

The process of comparing student responses to the DAPVU rubrics allowed me to determine when it would and when it would not be fitting to assign student understanding to levels in the DAPVU rubrics. It also helped me determine the types of evidence available for each of the knowledge dimensions. The DAPVU knowledge dimensions include facets of place value understanding and I was able to obtain evidence for four facets of place value understanding on the online DAPVU—Accuracy, Representation, Descriptive Language, and Depth.

I drew heavily on student responses and used the DAPVU rubric levels as a guide to code the student responses into evidence markers for each of the facets. My evidence markers are categorized descriptions of student responses that provide evidence of a student's understanding of various facets of place value. Most, but not all, facets have ordinal evidence markers. Each facet has at least one marker that signifies insufficient evidence of a student's understanding of that particular facet. The four facets of place value understanding do not represent an exhaustive list of all of the characteristics that make up place value understanding and I am merely using them as tools to describe the evidence of place value understanding I was able to find in student responses to the online DAPVU. I broke down each DAPVU problem situation by operation or structure, quantitative representation, and evidence marker in Table 4. Below Table 4, I briefly describe the Accuracy, Representation, Descriptive Language, and Depth facets and their related evidence markers. Appendix O contains a complete description of how to assign evidence markers to student responses on the five online DAPVU problem situations.

Table 4

Operation or Structure, Quantitative Representation, and Evidence Markers of DAPVU

DAPVU Problem Situations		Operation or Structure	Quantitative Representation	Evidence Markers			
				A	R	L	D
PS1	Pat's Skiing Competition	Addition	Familiar-Nonsystematic	X	X		X
PS2	Bobby's Squares	Notational	Familiar-Systematic			X	X
PS3	Chocolate Factory	Subtraction	Unfamiliar-Systematic	X	X		X
PS4	Rugolian Rug Merchant	Addition	Unfamiliar-Nonsystematic	X	X		X
PS5	Maria's Error Pattern	Subtraction	Familiar-Systematic	X		X	X

Note. The associated Evidence Markers are: Accuracy (A), Representation (R), Descriptive Language (L), and Depth (D).

Accuracy (Accurate Computation or Accurate Reproduction of Error)

I used the DAPVU's Accurate Computation (AC) and Error Reproduction (ER) rubrics to create the evidence markers for the Accuracy facet. Unlike the levels in the AC and ER rubrics that go beyond accuracy and include reasons for inaccuracy (i.e., computational error, conceptual error), the Accuracy evidence markers are based strictly on correctness. I do not believe sufficient information may be gleaned from the online DAPVU to make definite judgments about why students' responses are correct or incorrect.

Student responses to PS1, PS3, PS4, and PS5 may be categorized using the Accuracy facet. Responses to PS2 (Bobby's Squares) cannot be categorized in terms of accuracy because there are no computational responses to define as correct or incorrect. For PS1, PS3, and PS4, accuracy involves whether or not a student correctly solved computations. For accuracy evidence on these problem situations, I placed students in one of three evidence marker categories: “no computation,” “inaccurate computation,” or “accurate computation.”

For PS5 (Maria’s Error Pattern), the Accuracy evidence maker categories are more complicated. If Maria’s error pattern is adhered to, the first subtraction problem should have the correct answer (212) and the second and third subtraction problems should have specific incorrect answers (356 and 266). A student’s fully correct reproduction of the pattern would be: 212, 356, and 266. A partial reproduction of the pattern is indicated by only one of these answers being different from 212, 356, or 266. If the student’s first answer is different from 212 and the second and third answers are correct, it is likely (based on explanations from students) that the student recognized a pattern similar to, but different from Maria’s pattern. If the first answer is correct and only one of the second or third answers is different from 356 or 266, it is likely the student made a calculation error or has applied a pattern similar to, but different from Maria’s pattern. If answers to both of the second and third problems are different from 356 and 266, the student has not replicated the error pattern. Such students either used correct subtraction procedures or created random errors. Some students who did not replicate the errors provided correct or partially correct explanations about Maria’s error pattern. So, lack of error reproduction does not explicitly indicate that the student does not recognize the error pattern. For Accuracy on PS5, I placed students in one of three categories: “no reproduction of pattern,” “partial reproduction of pattern,” or “accurate reproduction of pattern.”

Representation

I created the Representation facet to distinguish between the ways students represent quantities in their responses. The Representation markers are not ordinal because students are not asked to provide the most efficient representations. The markers indicate when students represent quantities with digits and units (e.g., 5 minutes, 44 seconds, 21 milliseconds) and when they use a symbolic representation (e.g., 5:44.21). The Representation facet provides insight into what seems to be the most natural representation of quantity for students in these novel situations, and a student's choice of representation may signify whether or not the student recognizes the utility of symbolic representations in uncommon or mixed-radix place value structures. It was loosely informed by the Analysis of Symbolic Representation rubric. Only student responses to PS1, PS3, and PS4 may be categorized in terms of their choice of representation. Responses to PS2 and PS5 cannot be categorized in terms of the student's choice of representation because there are no computational responses in the former and students are requested to provide a specific type of representation in the latter.

Descriptive Language

The Descriptive Language evidence markers mirror the levels in the DAPVU'S Use of Descriptive Language rubrics. Student responses to PS2 and PS5 may be categorized in terms of the students' use of descriptive language because students are asked to describe concepts Bobby and Maria (hypothetical students in the PS2 and PS5) do and do not understand. The idea is find evidence that students are utilizing place value language in their descriptions, and the evidence markers describe increasing levels of sophistication in place value language use. Briefly, evidence at the lowest level may described as "no or inaccurate use of place value language" and evidence at the highest level may be described as "accurate, highly specific place value language." It is inappropriate to assign ordinal levels of place value understanding on PS1, PS3, and PS4 because students are asked how they solved those problem situations and are not probed as to why they chose a particular method.

Depth (Depth of Analysis or Depth of Understanding)

The Depth facet stems from a combination of the Analysis of Computation Method and Depth of Analysis rubrics. In problem situations in which respondents are to analyze the work of hypothetical students (PS2 and PS5), the Depth (of analysis) evidence markers mirror the DAPVU's Depth of Analysis levels. In other problem situations (PS1, PS3, and PS4), Depth (of understanding) is evidenced by students' descriptions of their methods and reflections about their problem-solving processes. I have included categorical markers that specify different reasons I couldn't find Depth evidence due to lack of information (e.g., estimation, insufficient evidence to decipher computation method). When students provide responses that are appropriate for Depth examination, they are placed into ordinal Depth marker categories. For PS2 and PS5, the markers range from "provides an analysis that is irrelevant, incorrect, or uninformative" to "develops an accurate and elaborate analysis of the place value concepts no understood by the child."

PS1, PS3, and PS4 Depth markers focus on evidence of understanding of the place value structure in the given problem. For example, Depth markers for PS3 range from "no evidence that the base-four place value structure is recognized and utilized" to "evidence that the base-four place value structure is utilized." Essentially, Depth, for all problem situations ranges from "no evidence of understanding" (with enough information to gauge depth) to "evidence of conceptual understanding."

Semi-Structured Interviews

Interviewees

I invited all participants that completed the surveys and the DAPVU to participate in a one-hour, one-one-one semi-structured interview at the end of the semester. I recorded the interviews with an audio-recorder and the interviewees solved math problems using a Livescribe Echo smartpen and pad. Four students agreed to an interview, but one student was unable to meet due to an emergency. I use pseudonyms to describe the three interviewees. I interviewed two female respondents, Mia and Casie, aged 25-29, from College B. They had the same Foundations course instructor, but were in different course sections. Mia self-reported as Hispanic, Latino or Spanish. She was enrolled part-time and working 40 or more hours per week. According to the demographic survey, she has one child, her mother received a high school diploma or GED, and Mia was not a first generation college student. Casie reported as Other/Multiracial on the demographic survey. She was not employed, but was going to school full-time and does a lot of babysitting. She has no children, is a first generation college student, and was unsure of her mother's level of education. I also interviewed Robert, a White, non-Hispanic male, age 30 to 39, who was enrolled part-time at College A. Robert has one child and was working 20-25 hours per week. His mother had some postsecondary education; he was not a first generation college student. All three interviewees were native English speakers.

Overview of Interviews

Interviews tend to be categorized by the degree to which the interview is systematized: unstructured, semi-structured, or structured. Semi-structured interviews are sometimes called semistandardized, loosely structured, guided-semistructured, or focused interviews (Berg, 2000; Bogdan & Biklen, 2003; Britten, 1995). These popular interviews fall somewhere in between unstructured and structured on the interview spectrum, so the control of the interview is more balanced between the interviewer and interviewee. At the outset, the interviewer asks open-ended questions that are related to the research question (Ginsburg, 1981). Many researchers create a list of possible questions to ask, but the interview is not scripted and diverging from the list is acceptable (Britten, 1995). During a semi-structured interview, the researcher does not function under the belief that questions on a predetermined list (created based on literature or what the researcher assumes to have the greatest importance) are the items the respondents see as central; hence, the researcher may follow an alternate line of reasoning introduced by the subject. According to Bogdan and Biklen (2003), "even when an interview guide is employed, qualitative interviews offer the interviewer considerable latitude to pursue a range of topics and offer the subject a chance to shape the content of the interviews" (pp. 94-95). While flexibly allowing the subject to help guide the interview, the researcher must also keep his or her own research question(s) in focus to make certain to obtain the needed data and not move too far off track (Spradley, 1979).

Jean Piaget is credited with inspiring clinical interview methodology, and his use of open-ended questions during interviews in the 1920s centered on uncovering children's knowledge structures and reasoning processes by studying their language use (R. Campbell, 2006; Clement, 2000; Elkind, 1964; Ginsburg, 1981, 1997; Lee, Russ, & Sherin, 2008). Over time, Piaget shifted from interviews with open-ended questions to observing children's solution strategies while they worked on concrete tasks and following up with questions (R. Campbell, 2006). Since the mid-1970s there have been extensions of his methods, including cognitive interviews, open-ended interviews, and think-aloud protocols (Clement, 2000). Clinical interviews are now used in psychological research on the mathematical thinking of people of all ages to discover cognitive processes, identify cognitive processes, and evaluate competence (Ginsburg, 1981).

Role and Responsibilities of the Interviewer

The interviewer must be cognizant of how his or her beliefs and background interplay with the interview and analysis processes (Corbin & Strauss, 2008). The researcher must also be aware of how he or she is viewed by the respondent and the effects of characteristics such as class, race, sex, and emotional nature of the topic; and the researcher can make appropriate adjustments in interviewing and data analysis techniques (Britten, 1995). Bogdan and Biklen (2003) claim, "Good interviewers communicate personal interest and attention to subjects by being attentive, nodding their heads, and using appropriate facial expressions to communicate" (p. 96). The researcher is to remain nonjudgmental and impartial. The interviewer should use probing questions that allow the interviewee to be placed in an expert role and insure deeper conversation (Bogdan & Biklen, 2003; Ginsburg, 1981; Spradley, 1979).

In his seminal publication about clinical interviews, Ginsburg (1997) provides a list of guidelines for conducting clinical interviews with children that are relevant to subjects of varying ages. Summarily: The interviewer should not discourage the subject's way of solving problems. The interviewer should refrain from making judgments about correctness—there may be correct reasoning behind a wrong answer or the subject is answering a different question than the interviewer intended, and there could be incorrect reasoning behind correct answers. It's important to ask fundamental questions, such as: "How did you figure that out?" or "Can you show me how you did it?" An alternative to directly asking the interviewee how he or she solved something, the interviewer may use reflection or echo the interviewee's response to elicit further information. The interviewer should ask for justification and follow the subject's line of reasoning, even if it runs contrary to what the interviewer expected at the outset. Related to this, the interviewer should not ask leading questions or correct the interviewee (unless it is a clinical teaching interview). The interviewer must respond to correct and incorrect answers in the same, non-judgmental way so the responsibility of determining correctness rests with the interviewee. The interviewer should be observant of non-verbal behaviors, not just verbal responses. Perhaps the most difficult guideline to follow is: Don't talk too much. The interviewer's role is to observe, probe, and listen. These guidelines are easiest to implement if one keeps in mind: the researcher is meant to learn or uncover the subject's understanding or thoughts about some topic (not the other way around).

Preparing for the Interviews

I created a loose template of topics and questions to direct my one-one-one, audio-recorded, semi-structured interviews. The interviews were set to last one hour. Because I was following a semi-structured interview format, I did not know how long it would take to get through each topic. As such, I tried to address the most crucial items near the beginning of the interview. After asking about their background and general experiences in their Foundations course, I asked more specific questions related to the goals of this study. I used the background and general experiences questions to elicit responses about what the interviewees believed to be the most salient features of the curriculum, whether or not it seemed different from prior math courses, and whether or not they thought it had a positive impact. I used "grand tour questions" (Leech, 2002), such as, "Can you describe a typical day [in your Foundations class]?" The more specific questions were meant to ascertain whether or not they felt they had changed with respect to their math anxiety, math mindset, math self-efficacy, math belonging, and college belonging. I addressed issues of persistence, including their reactions to math failure, and their beliefs about their ability to transfer knowledge or mindsets gained in the Foundations course to subsequent courses and their mathematical lives. All of these items make up the core of what I desired to gain from the interviews because they address my research questions in general—Did their dm-noncognitive factors change? Do they see themselves as able to productively persist and transfer their knowledge? Do they attribute changes to the Foundations course? Do changes measured by the pre-post survey or the post-then survey more accurately represent their beliefs?

I included additional, time-permitting topics in the interview template: I integrated a rough think-aloud with verbal probing for some DAPVU problem situations they solved online and similar APVU problem situations. I used prompts to elicit reflections on their problem-solving approach and reasons for persistence (or lack thereof) in novel situations. I also included questions about the format of the DAPVU to ascertain its appropriateness as a measure of their preparation for future learning. I asked logistical questions about the in-class surveys; specifically, I asked if they had already covered any of the survey topics prior to completing the pre-survey and I asked questions to determine the clarity of the post-then survey format. I asked the former survey question because I was suspicious that some students took the pre-survey after the mindset lesson and I asked the latter question to determine if post-then survey results were contaminated because of respondent confusion. The final questions in the interview relate to the student's course success with respect to grades. The interview template, including the clinical interview problems, is in Appendix P.

Beginning the Interview

Before the interview I briefly described the study and interview. I then asked interviewees to verbalize their consent to be audio-recorded, even though I had obtained their written consent. I explained the audio-recordings would not be shared with others and the recordings would allow me to focus on the interview instead of having to take extensive notes. I also said that I wanted to be sure to represent them accurately and an audio recording would better serve this purpose than my memory. I assured them that their responses in the interview and other parts of the study are completely confidential. I told them that I had not looked at their responses to any of my study instruments because I was going to try not to ask any leading questions and I told them not to worry about whether their responses to the instruments matched what they said in the interview because I could get a more accurate picture of what they think in person than on an instrument. When respondents asked questions that would muddy results, I explained that I could not discuss the specifics of my study because it may impact their responses, but I would be happy to provide more details after the interview.

I attempted to put the interviewees at ease and in an expert role. I explained that I did not create the Foundations course, but I am interested in providing the curriculum creators with information that could help the designers revise the course and better serve the needs of students. I emphasized that I was judging the course, not them or their instructors, and I was interested in their perceptions of the course so they could help me with this goal and benefit future students taking the Foundations course. I reiterated that I desired their bluntness and honesty because nobody would be better at critiquing the course than the students themselves. I also explained that I would ask for feedback on my study instruments if we had time so that I could improve the instruments for future research. Again, I explained that they were the most appropriate people to ask these questions because they were the ones who completed the instruments and I wanted their critical feedback to help me improve as a researcher.

I reiterated that I was purely interested in how they were thinking about the problem situations during the think-aloud portion of the interview, and it was extremely helpful for them to verbalize their thoughts. For example: “I don’t want to make assumptions based on my own ideas, but I can’t see inside your brain. I know it can be really hard to explain thinking, especially while you are solving a problem, but please try to talk about what you are doing and why as much as you can. I may ask a bunch of questions that seem redundant, but I really want to make sure I capture what you are thinking and not what I think you are thinking.” I explained how difficult it is to understand how people are reasoning using just the responses to the online DAPVU, so any insights they could share would be very valuable. I noted that I had never administered the assessment before and I wanted to get feedback that could help me make it better. Again, this was an attempt to put the interviewee in an expert role. I tried to follow Ginsburg’s guidelines by asking probing, non-leading questions and refraining from the tendency to correct and teach, even when they asked specifically if they were correct. I discuss interview results in Chapter 4.

DATA COLLECTION PROCEDURES

Participants completed pre-surveys, post-then-surveys, and a place value assessment of transfer (DAPVU). I obtained participants’ final exam grades, final course grades, and attendance records (when available). I conducted one-on-one, audio-recorded, semi-structured interviews with a small subset of participants to provide additional insight not gleamed from the other measures.

At the beginning of the semester, I provided packets to instructors who agreed to assist with the study. The packets contained a statement explaining the study, a consent description, scantrons, and pre-surveys. In the second or third week of class, the instructors passed out the consent description and pre-survey, and described the study using a statement I provided (see Appendix D). The instructors provided an overview of the study, pointed out features of the consent description (e.g., participation is voluntary, how to contact me, how much time the study will take), and explained the compensation participants would receive (HEB gift cards in varying amounts based on extent of study participation). All students were allowed to keep a copy of the consent description. Students who wished to participate signed the consent form that is on the first page of the pre-survey. Students were given five minutes of class time to complete the pre-survey on scantrons. After completing the consent form and pre-survey, students placed the consent forms and surveys in a provided packet so the instructor would not be privy to the names or responses of the participants. The instructor gave the packets to a designated person at the college who then gave the packets to me.

Just after the midpoint of the semester, I sent packets to instructors who agreed to assist. The packets contained a statement that reminds students of the study (see Appendix F), additional consent descriptions, post-then-surveys with a consent form on the first page, and scantrons. A few weeks before the end of the semester, the instructors provided another brief description of the study using the provided statement (see Appendix F) and passed out the post-then-survey. The post-then-survey has three sections in this order: post, then, and demographic questions. Students were instructed not to begin the then-survey until they completed the post-survey and they were instructed not to begin the demographic section until they completed the then-survey. Students were given ten minutes of class time to complete the post-then-survey on scantrons. All students who wanted to participate were allowed to do so, even if they did not participate at the beginning of the semester. Students who had not previously given their consent were allowed to do so at this time by filling out the consent form on the first page of the survey. After completing the consent form and post-then-survey, students placed the consent forms, scantrons, and surveys in a provided packet so the instructor would not be privy to the names or responses of the participants. The instructor gave the packets to a designated person at the college who then gave the packets to me.

Students who signed the consent forms and completed the surveys were contacted via email with a link to the DAPVU. The DAPVU was administered online through Qualtrics software (Qualtrics, Provo, UT). The students took the AVPU-D at the end of the semester only. There are two main reasons for giving the assessment at one time point: 1) The Foundations curriculum was designed so students first encounter success, not failure. If they were to take a pretest on which they did not feel successful, this would run counter to one of the main curriculum goals. 2) Pretesting provides opportunities for learning from the test and this introduces the risk of tainting the validity of posttest results (D. Campbell & Stanley, 1963; Rudestam & Newton, 1992). All students who completed both surveys, completed the online DAPVU, and consented to interviews were contacted for one-on-one, audio-recorded interviews with me.

METHODS OF DATA ANALYSIS

In this study, I attempted to answer four research questions. The key dm-noncognitive factor variables in my models are: math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness. The semester outcome variables in my models are: numeric math course grade, numeric math final exam grade, and math course percent attendance. The demographic control variables are: gender, age, race/ethnicity, enrollment status, job status, number of dependents, native language, maternal education, and first generation college student. The DAPVU and one-on-one semi-structured interviews were used as qualitative measures.

I used multilevel models to answer my quantitative research questions, while accounting for nesting (e.g., time points nested within students and students nested within classes). Multilevel models can evaluate relationships across different levels (e.g., when students are at level-1 and class section is at level-2) while parsing the effects of between- and within-group variance (Raudenbush & Bryk, 2002). Woltman, Feldstain, MacKay, and Rocchi (2012) note many advantages to using multilevel models: Multilevel models "can accommodate non-independence of observations, a lack of sphericity, missing data (at level-1), small and/or discrepant group sample sizes, and heterogeneity of variance across repeated measures" (p. 56). Plus, they do not distort effect size estimates or standard errors. Multilevel models need large sample sizes to detect effects, especially at level-1, and power is improved by increasing the number of groups more so than by increasing the number of observations per group (Woltman et al., 2012). I conducted a power analysis, using G*Power 3.1.9.2 (Faul, Erdfelder, Buchner, & Lang, 2009), to determine the number of participants I needed to achieve a power level of .8, while setting the significance to the alpha level needed for each research question, and my sample size was adequate for each multilevel model. I discuss the results of the power analyses in the results sections for each research question. I used IBM SPSS Statistics (Version 23.0) for my quantitative data analyses.

Questions 1A and 1B: Changes in DM-Noncognitive Factors

Do students exhibit differences over time in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness?

Question 1A: Pre-Survey vs. Post-Survey

Do students exhibit beginning-of-semester to end-of-semester differences in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness?

RQ1A Research Hypothesis: Students will exhibit beginning-of-semester to end-of-semester improvements in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness.

Question 1B: Pre-Survey vs. Then-Survey

Do students exhibit pre-survey to then-survey differences in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness?

RQ1B Research Hypothesis: Students will exhibit pre-survey to then-survey differences in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness.

To address these research questions, I ran a separate multilevel model for each the following dependent variables: math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness. The models included a fixed effect of the within-subjects variable, time. My models also included fixed effects for school and the following demographic variables: gender, age, race/ethnicity, enrollment status, job status, number of dependents, native language, maternal education, and first generation college student. The models included random intercepts for participant and section. I used the comparison between pre-survey and post-survey scores to answer Research Question

1A[34] and the comparison between pre-survey and then-survey scores to answer Research Question 1B. I used an alpha level of .01 in all statistical tests for Research Questions 1A and 1B to account for multiple comparisons.

<u>Multilevel Models for Research Questions 1A and 1B</u>

The model below was used to test whether or not students' reports of math equanimity changed over time. The other four models are parallel to this model, with math equanimity replaced by math mindset, math self-efficacy, math belongingness, or college belongingness.

Level-1 Equation

$$\mathrm{Equanimity}_{\mathrm{hij}} = \beta_{0\mathrm{ij}} + \beta_1 * (\mathrm{Time}_{\mathrm{Post}})_{\mathrm{hij}} + \beta_2 * (\mathrm{Time}_{\mathrm{Then}})_{\mathrm{hij}} + \beta_3 * (\mathrm{School})_{\mathrm{hij}} + \beta_4 * (\mathrm{Gender})_{\mathrm{hij}} + \beta_5 * (\mathrm{Enrollment})_{\mathrm{hij}} + \beta_6 * (\mathrm{Kids})_{\mathrm{hij}} + \beta_7 * (\mathrm{English})_{\mathrm{hij}} + \beta_8 * (\mathrm{FirstGen})_{\mathrm{hij}} + \beta_9 * (\mathrm{Race}_{\mathrm{Black}})_{\mathrm{hij}} + \beta_{10} * \left(\mathrm{Race}_{\mathrm{Hispanic}}\right)_{\mathrm{hij}} + \beta_{11} * (\mathrm{Race}_{\mathrm{White}})_{\mathrm{hij}} + \beta_{12} * (\mathrm{Age})_{\mathrm{hij}} + \beta_{13} * (\mathrm{Job})_{\mathrm{hij}} + \beta_{14} * (\mathrm{MaternalEd})_{\mathrm{hij}} + \varepsilon_{\mathrm{hij}}$$

In the above equation, $\mathrm{Equanimity}_{\mathrm{hij}}$ is the math equanimity score reported at the h^{th} time by the i^{th} student in the j^{th} section.

Level-2 Equation

$$\beta_{0\mathrm{ij}} = \gamma_{00\mathrm{j}} + \mu_{0\mathrm{ij}}$$

In the above equation, μ_{0ij} is a normally distributed variable with a mean of zero ($\mu_{0ij} \sim N[0, {\theta_{\mu_j}}^2]$).

[34] Considering my literature review discussion that then-surveys are more accurate than pre-surveys, it may seem counterintuitive that I would use a pre-post-survey comparison to address Research Question 1A. Even though retrospective pre-surveys have been lauded in some fields as more meaningful than traditional pre-surveys, they are not widely used in math education research—In fact, most education researchers with whom I discussed my study were completely unaware of the existence of thentests. I decided a priori to use change scores between pre- and post-surveys to address Research Question 1A because I did not want my results questioned solely based on my using a measure that is not currently widely accepted, and Research Question 1B affords me the chance to compare traditional and retrospective pre-surveys.

Level-3 Equation

$\gamma_{00j} = \tau_{000} + \alpha_{00j}$

In the above equation, α_{00j} is a normally distributed variable with a mean of zero ($\alpha_{00j} \sim N[0, {\theta_\alpha}^2]$).

Question 2: Outcomes and Changes in DM-Noncognitive Factors

Do beginning-of-semester to end-of-semester differences in students' math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness predict semester outcomes[35] (math course grade, math final exam grade, and math course percent attendance)?

RQ2 Research Hypothesis: Beginning-of-semester to end-of-semester differences in students' math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness will predict semester outcomes.

To address this research question, I ran three multilevel models, one for each of my dependent variables (math course grade, math final exam grade, and math course percent attendance). I used the change scores in dm-noncognitive factors (post-survey minus pre-survey for each of: math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness) as my independent variables, and I treated section as a random effect. My models control for pre-survey scores and include fixed effects for school and the following demographic variables: gender, age, race/ethnicity, enrollment status, job status, number of dependents, native language, maternal education, and first generation college student. I used an alpha level of .017 in all statistical tests for Research Question 2 to account for multiple comparisons.

[35] I was unable to use Developmental Assessment of Place Value Understanding scores as a quantitative outcome measure.

Multilevel Models for Research Question 2

The model below was used to test whether or not changes in students' reports of dm-noncognitive factors—math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness—are associated with their class grade. The other two models are parallel to this model, with Course Grade replaced by Final Exam or Attendance.

Level-1 Equation

$$\begin{aligned}\text{Course Grade}_{\text{ij}} = {} & \beta_{0\text{j}} + \beta_1 * (\text{Equanimity}_{\text{Post-pre}})_{\text{ij}} + \beta_2 * (\text{Mindset}_{\text{Post-Pre}})_{\text{ij}} + \beta_3 * \\ & (\text{SelfEfficacy}_{\text{Post-Pre}})_{\text{ij}} + \beta_4 * (\text{MathBelonging}_{\text{Post-Pre}})_{\text{ij}} + \beta_5 * \\ & (\text{CollegeBelonging}_{\text{Post-Pre}})_{\text{ij}} + \beta_6 * (\text{School})_{\text{ij}} + \beta_7 * (\text{Gender})_{\text{ij}} + \beta_8 * \\ & (\text{Enrollment})_{\text{ij}} + \beta_9 * (\text{Kids})_{\text{ij}} + \beta_{10} * (\text{English})_{\text{ij}} + \beta_{11} * (\text{FirstGen})_{\text{ij}} + \beta_{12} * \\ & (\text{Race}_{\text{Black}})_{\text{ij}} + \beta_{13} * (\text{Race}_{\text{Hispanic}})_{\text{ij}} + \beta_{14} * (\text{Race}_{\text{White}})_{\text{ij}} + \beta_{15} * (\text{Age})_{\text{ij}} + \\ & \beta_{16} * (\text{Job})_{\text{ij}} + \beta_{17} * (\text{MaternalEd})_{\text{ij}} + \beta_{18} * (\text{Equanimity}_{\text{Pre}})_{\text{ij}} + \beta_{19} * \\ & (\text{Mindset}_{\text{Pre}})_{\text{ij}} + \beta_{20} * (\text{SelfEfficacy}_{\text{Pre}})_{\text{ij}} + \beta_{21} * (\text{MathBelonging}_{\text{Pre}})_{\text{ij}} + \beta_{22} * \\ & (\text{CollegeBelonging}_{\text{Pre}})_{\text{ij}} + \varepsilon_{\text{ij}}\end{aligned}$$

In the above equation, $\text{Course Grade}_{\text{ij}}$ is the class grade for the i^{th} student in the j^{th} section.

Level-2 Equation

$$\beta_{0\text{j}} = \gamma_{00} + \mu_{0\text{j}}$$

In the above equation, μ_{0j} is a normally distributed variable with a mean of zero ($\mu_{0j} \sim N[0, {\theta_\mu}^2]$).

Question 3: Evidence of Place Value Concept Transfer

Do students exhibit evidence of their ability to transfer their knowledge to novel place value problems?

RQ3 Research Hypothesis: Students will exhibit evidence of their ability to transfer their knowledge to novel place value problems.

To address this question, I looked for evidence of procedural and conceptual understanding by qualitatively analyzing student responses to the online DAPVU. I used this information and the DAPVU knowledge dimension rubrics to categorize student responses in terms of four facets of place value understanding—Accuracy, Representation, Depth, and Descriptive Language. I used evidence markers from the Accuracy, Depth (of understanding), and Representation facets to categorize the place value understanding evidenced by students on Problem Situation 1 (Pat's Skiing Competition), Problem Situation 3 (Chocolate Factory), and Problem Situation 4 (Rugolian Rug Merchant). I categorized evidence of students' place value understanding on Problem Situation 2 (Bobby's Squares) in terms of descriptive language and depth (of analysis). I used evidence markers from the Accuracy, Depth (of analysis), and Descriptive Language facets to categorize the students' understanding of Problem Situation 5 (Maria's Error Pattern).

DOCUMENT ROADMAP

In Chapter 3, I listed my research questions and hypotheses. I used results from my demographic survey to describe my participants. I detailed my data collection measures: pre-, post-, and then-surveys of dm-noncognitive factors (math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness); math course outcomes (math course grade, math final exam grade, and math course attendance); the Developmental Assessment of Place Value Understanding; and semi-structured, one-on-one interviews. I described how I collected and planned to analyze my data with respect to my research questions. I present the results of my data analyses in Chapter 4.

Chapter 4: Results

Key Variables

DM-Noncognitive Factor Variables

The key dm-noncognitive factor variables in my models for Research Questions 1A, 1B, and 2 are measured via pre-post-then-surveys with Likert-type[36] items and are treated as continuous variables because there are 5 or more ordinal response levels for each item. Scores for math equanimity, math self-efficacy, math belongingness, and college belongingness range from 0 to 4, with 4 being more indicative of a higher dm-noncognitive factor. Scores for math mindset range from 0 to 5, with 5 being more indicative of a higher dm-noncognitive factor.

Semester Outcome Variables

The semester outcome variables in my models for Research Question 2 are math course grade, math final exam grade, and math course percent attendance. These variables are continuous and measured on a scale from 0 to 100.

Demographic Control Variables

The dichotomous demographic variables in my models for Research Questions 1A, 1B, and 2 are: gender (male or female), enrollment status (part time or full time), number of dependents (none or at least one)[37], English as native language (yes or no), and first generation college student (yes or no). Because some ethnic/racial groups were considerably underrepresented, I collapsed the ethnicity/race variable to four categories:

---

[36] According to Uebersax (2006), the items on the five measures are not Likert items in the strictest sense because the verbal labels are not all "bivalent and symmetrical about a neutral middle. They meet the requirements for Likert-type items, with the exception of response levels being "anchored with consecutive integers". For use with Scantrons, I anchored the response levels with consecutive letters.

[37] I collapsed results for "Number of Dependents" from four categories to two to simplify my models.

Hispanic, Latino, or Spanish; Black or African American, Non-Hispanic; White, Non-Hispanic; and Multi-Racial or Other. Because ordinal, categorical variables with five or more categories may be treated as continuous, I treated maternal education[38], age, and job status as continuous variables (The categories are listed in the Demographics section and in Appendix I).

Qualitative Measures

The online DAPVU is a qualitative measure for Research Question 3. The one-on-one semi-structured interviews served to provide more insight into results from Research Questions 1A, 1B, 2, and 3.

ASSUMPTIONS

The assumptions of multilevel modeling are the same as the assumptions of multiple regression (except for independence within groups): (a) the residuals (errors) are normally distributed; (b) there is homoscedasticity (homogeneity of variance) of the residuals; (c) there is a linear relationship between the predictor variables and the dependent variable; (d) there is no multicollinearity[39]; and (e) there are no influential outliers. My data met the assumptions for multilevel modeling with two exceptions that violated the normality assumption.

First, the distribution of the final exam variable for Research Question 2 was bimodal, with one mode at the peak of a normal curve and one mode at 0 (high frequency of zeros). The distribution suggests two different groups are being represented by the data—students who took the exam and students who did not take the exam. While there is

[38] Nineteen students chose "Unknown" for Maternal Education. I treated those responses as missing data to make Maternal Education an ordinal variable.

[39] See Appendix J for correlation matrices of the dm-noncognitive factor scores on the pre-, post-, and then-surveys.

a chance a small number of students took the final exam and legitimately received a 0, there is no way to distinguish those students from students who did not take the exam. To meet the normality assumption, I excluded students who received a final exam grade of 0 (n=59) from all analyses that included the final exam variable.

Second, the attendance variable for Research Question 2 was negatively skewed. I transformed the variable by reflecting it and then taking the natural logarithm. The transformed attendance variable follows a normal distribution. I ran the models with the original attendance variable and again with the transformed attendance variable, and the interpretation of the results was the same in both analyses. The only differences relate to the demographic variables (i.e., the variables for self-efficacy pre-survey score and number of dependents were only significant when using the transformed attendance variable). I report results for the original attendance variable in the results section for Research Question 2 and report results for the transformed attendance variable in Appendix K.

QUESTIONS 1A AND 1B: CHANGES IN DM-NONCOGNITIVE FACTORS

Do students exhibit differences over time in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness?

Question 1A: Pre-Survey vs. Post-Survey

Do students exhibit beginning-of-semester to end-of-semester differences in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness?

> *RQ1A Research Hypothesis:* Students will exhibit beginning-of-semester to end-of-semester improvements in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness.

Question 1B: Pre-Survey vs. Then-Survey

Do students exhibit pre-survey to then-survey differences in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness?

RQ1B Research Hypothesis: Students will exhibit pre-survey to then-survey differences in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness.

To address these research questions, I ran a separate multilevel model for each of the following dependent variables: math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness. The models include a fixed effect of the within-subjects variable, time. The models also include fixed effects for school and the following demographic variables: gender, age, race/ethnicity, enrollment status, job status, number of dependents, native language, maternal education, and first generation college student. I based my significance level on a Bonferroni correction for multiple comparisons and set my alpha level to .01 for all analyses in Research Questions 1A and 1B. The initial models included random intercepts for participant and section. However, the random intercept for section was not significant for any of the models ($p > .01$), and I did not include section as a random effect in the final models. I used the comparison between pre-survey and post-survey scores to answer Research Question 1A and the comparison between pre-survey and then-survey scores to answer Research Question 1B.

It is not possible to conduct a power analysis for a mixed model with repeated measures using currently available software without both a treatment and control group. I used G*Power 3.1.9.2 to conduct a power analysis that most closely represented my models. I ran a power analysis for Research Questions 1A and 1B based on power analyses for repeated measures ANOVA. A sample size of 79 subjects is necessary to detect a medium effect size (f = .25) with a power of .80 for an alpha of .01 with 3 time points; my sample size was adequate to address Research Questions 1A and 1B.

<u>Final Multilevel Models for Research Questions 1A and 1B</u>

The model below was used to test whether or not students' reports of math equanimity changed over time. The other four models are parallel to this model, with math equanimity replaced by math mindset, math self-efficacy, math belongingness, or college belongingness.

Final Level-1 Equation

$$\mathrm{Equanimity}_{hi} = \beta_{0i} + \beta_1 * (\mathrm{Time}_{Post})_{hi} + \beta_2 * (\mathrm{Time}_{Then})_{hi} + \beta_3 * (\mathrm{School})_{hi} + \beta_4 * (\mathrm{Gender})_{hi} + \beta_5 * (\mathrm{Enrollment})_{hi} + \beta_6 * (\mathrm{Kids})_{hi} + \beta_7 * (\mathrm{English})_{hi} + \beta_8 * (\mathrm{FirstGen})_{hi} + \beta_9 * (\mathrm{Race}_{Black})_{hi} + \beta_{10} * \left(\mathrm{Race}_{Hispanic}\right)_{hi} + \beta_{11} * (\mathrm{Race}_{White})_{hi} + \beta_{12} * (\mathrm{Age})_{hi} + \beta_{13} * (\mathrm{Job})_{hi} + \beta_{14} * (\mathrm{MaternalEd})_{hi} + \varepsilon_{hi}$$

In the above equation, $\mathrm{Equanimity}_{hi}$ is the math equanimity score reported at the h^{th} time by the i^{th} student.

Final Level-2 Equation

$$\beta_{0i} = \gamma_{00} + \mu_{0i}$$

In the above equation, μ_{0i} is a normally distributed variable with a mean of zero ($\mu_{0i} \sim N[0, {\theta_\mu}^2]$).

Summary Results for Research Questions 1A and 1B

Table 5 lists results from the tests of fixed effects for each of the five models used to address Research Questions 1A and 1B. The table includes the significant and non-significant variables of interest (pre-, post-, and then-scores for math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness), as well as the significant control variables. Control variables that were not significant for a model are excluded from the results presented for that particular model, $p > .01$. The sections following Table 5 provide more information about the significant variables in each model.

Table 5

Tests of Fixed Effects for Variables of Interest and Significant Control Variables in Research Question 1A and 1B Models

Models	Variables	df_1, df_2	*F*	*Wald Z*	*p*
Math Equanimity					
	Time	2, 548.16	10.21		.000
	Random Intercept			9.18	.000
Math Mindset					
	Time	2, 545.43	2.82		>.01
	Random Intercept			9.62	.000
Math Self-Efficacy					
	Time	2, 542.73	1.75		>.01
	School	1, 277.28	9.39		.002
	Random Intercept			10.13	.000
Math Belonging					
	Time	2, 542.19	2.42		>.01
	School	1, 275.45	19.03		.000
	Native Language	1, 277.87	7.40		.007
	Age	1, 277.94	9.01		.003
	Random Intercept			9.17	.000
College Belonging					
	Time	2, 536.70	3.83		>.01
	School	1, 275.28	7.07		.008
	Age	1, 276.98	14.86		.000
	Random Intercept			9.80	.000

Note. This table includes results for both significant and non-significant variables of interest—pre-, post-, and then-scores for math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness—as well as significant control variables. For all Research Question 1A and 1B models, $n = 292$. Significance was set at $p < .01$.

Detailed Results for Question 1A

Do students exhibit beginning-of-semester to end-of-semester differences in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness?

> *RQ1A Research Hypothesis:* Students will exhibit beginning-of-semester to end-of-semester improvements in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness.

I used the comparison between pre-survey and post-survey scores to answer Research Question 1A. I used an alpha level of .01 in all statistical tests for Research Question 1A to account for multiple comparisons.

Math Equanimity (Pre vs. Post)

There was a significant effect of time on students' equanimity scores (F (2, 548.16) = 10.21, p = .000). Students were more math equanimous at the end of the semester (M = 2.26, SE = .11) than at the beginning (M = 2.07, SE = .11) of the semester, p = .002. Figure 1 contains a boxplot of the distribution of math equanimity scores on the pre- and post-survey. The random effect of student was significant (Wald Z = 9.18, p = .000). The other control variables were not significant, p > .01.

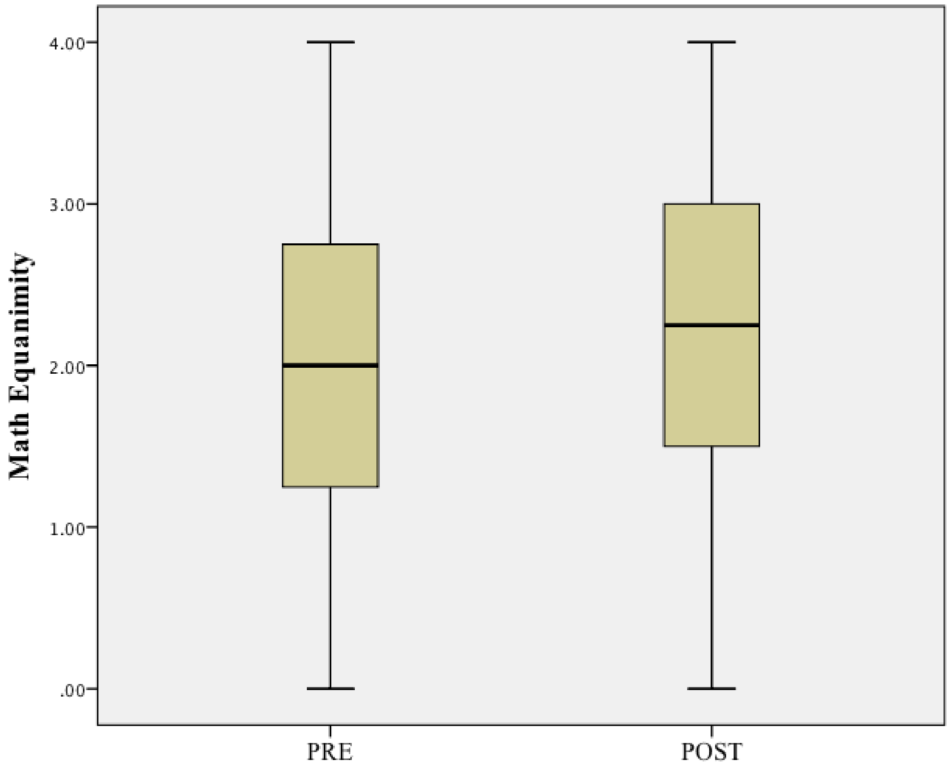


Figure 1. Distribution of pre- and post-survey math equanimity scores

Math Mindset (Pre vs. Post)

There was no effect of time on students' mindset scores (*F* (2, 545.43) = 2.82, $p > .01$), meaning students' math theory of intelligence did not change significantly from beginning ($M = 2.97$, $SE = .16$) to end ($M = 3.11$, $SE = .16$) of semester. Figure 2 contains a boxplot of the distribution of math mindset scores on the pre- and post-survey. The random effect of student was significant (Wald $Z = 9.62$, $p = .000$). The other control variables were not significant, $p > .01$.

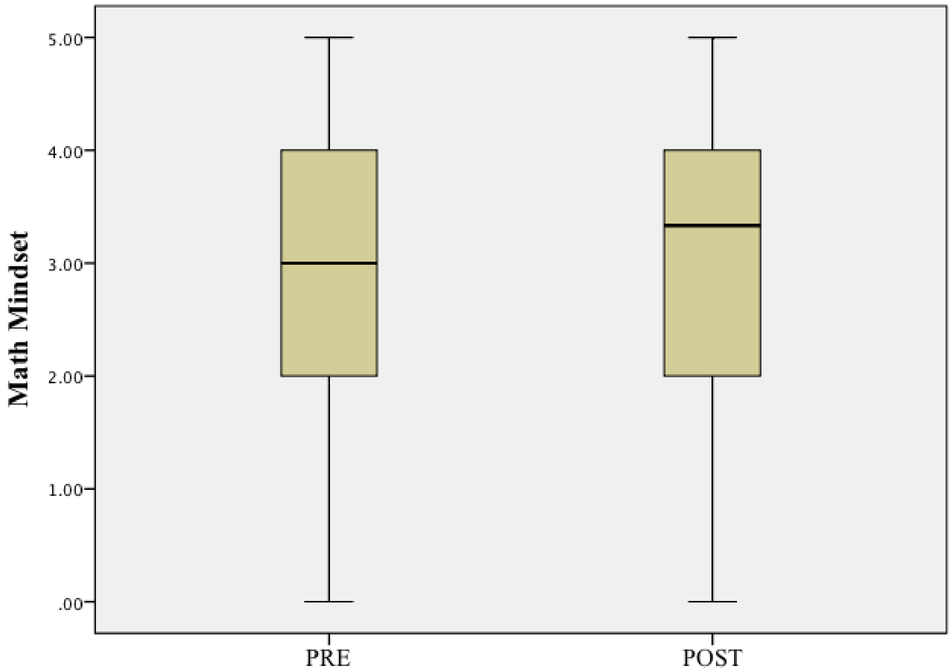


Figure 2. Distribution of pre- and post-survey math mindset scores

Math Self-Efficacy (Pre vs. Post)

There was not a significant effect of time on students' self-efficacy scores (F (2, 542.73) = 1.75, $p > .01$)), meaning students' math self-efficacy remained relatively stable from beginning (M = 2.62, SE = .13) to end (M = 2.65, SE = .13) of semester. Figure 3 contains a boxplot of the distribution of math self-efficacy scores on the pre- and post-survey. The fixed effect of school was significant (F (1, 277.28) = 9.39, p = .002), with students at College B (M = 2.84, SE = .14) reporting significantly higher self-efficacy scores than students at College A (M = 2.48, SE = .13). The random effect of student was significant (Wald Z = 10.13, p = .000). The other control variables were not significant, $p > .01$.

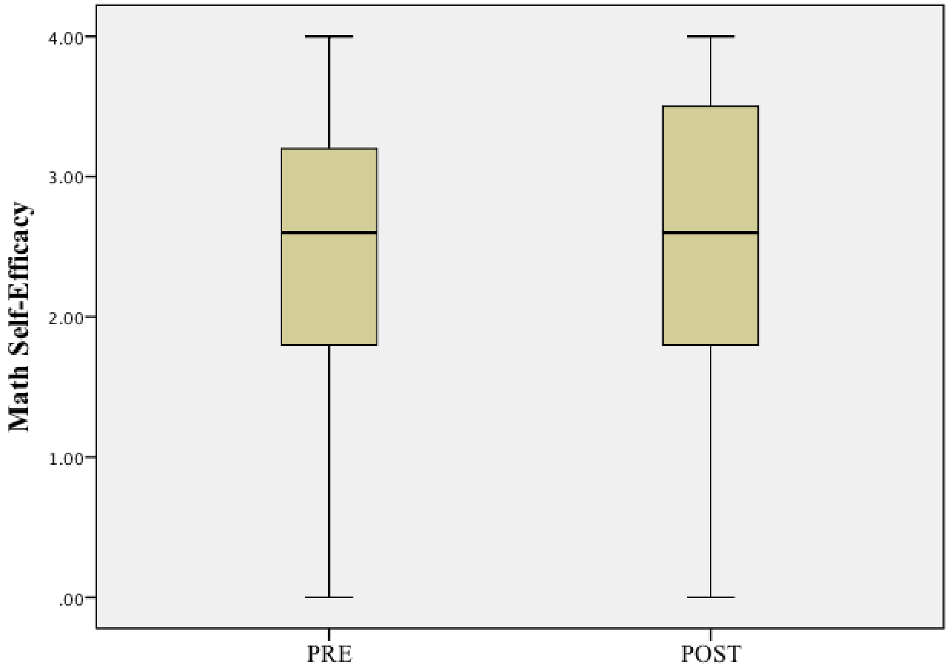


Figure 3. Distribution of pre- and post-survey math self-efficacy scores

Math Belongingness (Pre vs. Post)

There was not a significant effect of time on students' math belongingness scores (F (2, 542.19) = 2.42, p > .01)), meaning students' sense of math belonging remained relatively stable from beginning (M = 2.58, SE = .15) to end (M = 2.69, SE = .15) of semester. Figure 4 contains a boxplot of the distribution of math belongingness scores on the pre- and post-survey. The fixed effect of school was significant (F (1, 275.45) = 19.03, p = .000)), with students at College B (M = 2.96, SE = .16) reporting significantly higher math belongingness scores than students at College A (M = 2.38, SE = .15). The fixed effect of native language was significant (F (1, 277.87) = 7.40, p = .007)), with non-native English speakers (M = 3.00, SE = .25) reporting higher math belongingness scores than native English speakers (M = 2.34, SE = .09). The fixed effect of age was significant (F (1, 277.94) = 9.01, p = .003)); for every unit increase in age, a .13 increase in math belongingness score is expected. The random effect of student was significant (Wald $Z = 9.17$, $p = .000$). The other control variables were not significant, $p > .01$.

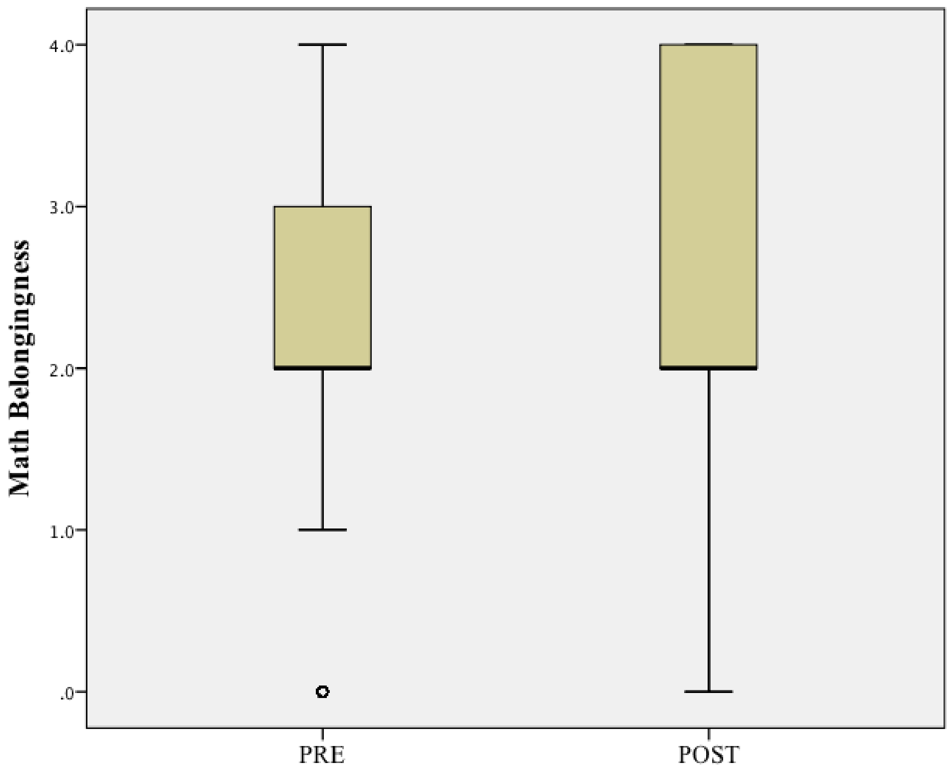


Figure 4. Distribution of pre- and post-survey math belongingness scores

College Belongingness (Pre vs. Post)

There was not a significant effect of time on students' college belongingness scores (F (2, 536.70) = 3.83, $p > .01$), meaning students' sense of college belonging remained relatively stable from beginning ($M = 2.99$, $SE = .15$) to end ($M = 2.87$, $SE = .15$) of semester. Figure 5 contains a boxplot of the distribution of college belongingness scores on the pre- and post-survey. The fixed effect of school was significant (F (1, 275.28) = 7.07, $p = .008$)), with students at College B ($M = 3.07$, $SE = .17$) reporting significantly higher college belongingness scores than students at College A (*M* = 2.71, *SE* = .15). The fixed effect of age was significant (*F* (1, 276.98) = 14.86, *p* = .000)); for every unit increase in age, a .17 increase in college belongingness score is expected. The random effect of student was significant (Wald *Z* = 9.80, *p* = .000). The other control variables were not significant, *p* > .01.

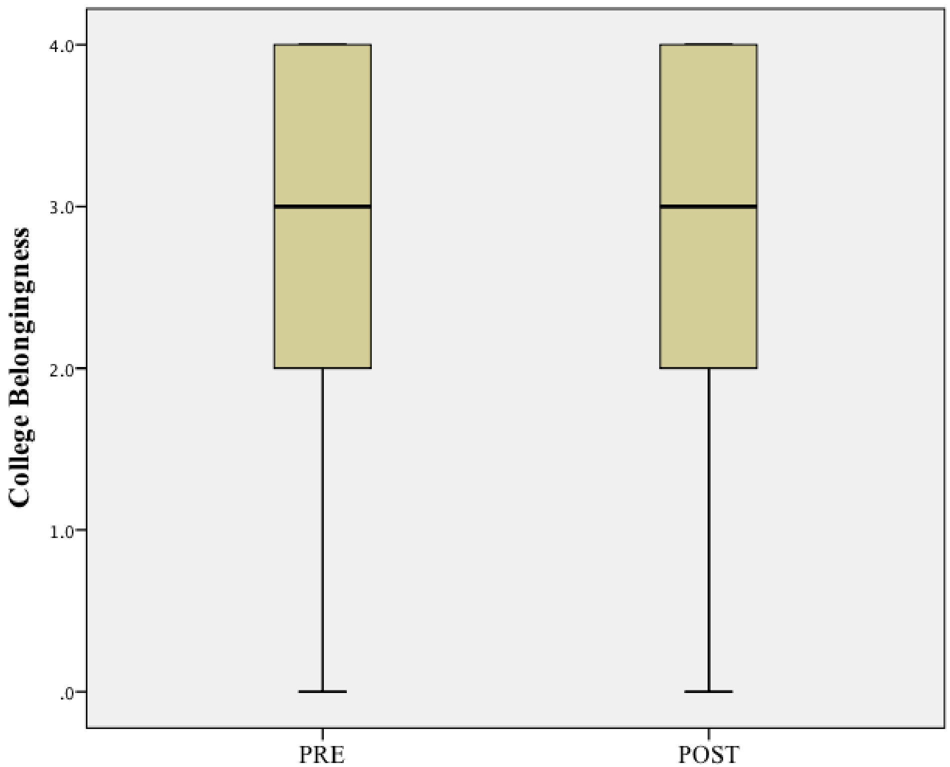


Figure 5. Distribution of pre- and post-survey college belongingness scores

Detailed Results for Question 1B

Do students exhibit pre-survey to then-survey differences in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness?

RQ1B Research Hypothesis: Students will exhibit pre-survey to then-survey differences in their math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness.

I reported the overall results of the multilevel models used to address Research Questions 1A and 1B in Table 5 above. I described the significant demographic variables in the multilevel models used to address Research Questions 1A and 1B in the detailed results section for Research Question 1A. In the following sections, I provide details about the differences between pre- and then-scores on the surveys, but I do not repeat the results pertaining to the demographic variables. I used an alpha level of .01 in all statistical tests for Research Question 1B to account for multiple comparisons.

Math Equanimity (Pre vs. Then)

As I reported in the results of Research Question 1A, there was a significant effect of time on students' equanimity scores (F (2, 548.16) = 10.21, p = .000). Students exhibited response shift, reporting greater math equanimity on the then-survey (M = 2.31, SE = .11) than the pre-survey (M = 2.07, SE = .11), p = .000. Figure 6 contains a boxplot of the distribution of math equanimity scores on the pre- and then-survey.

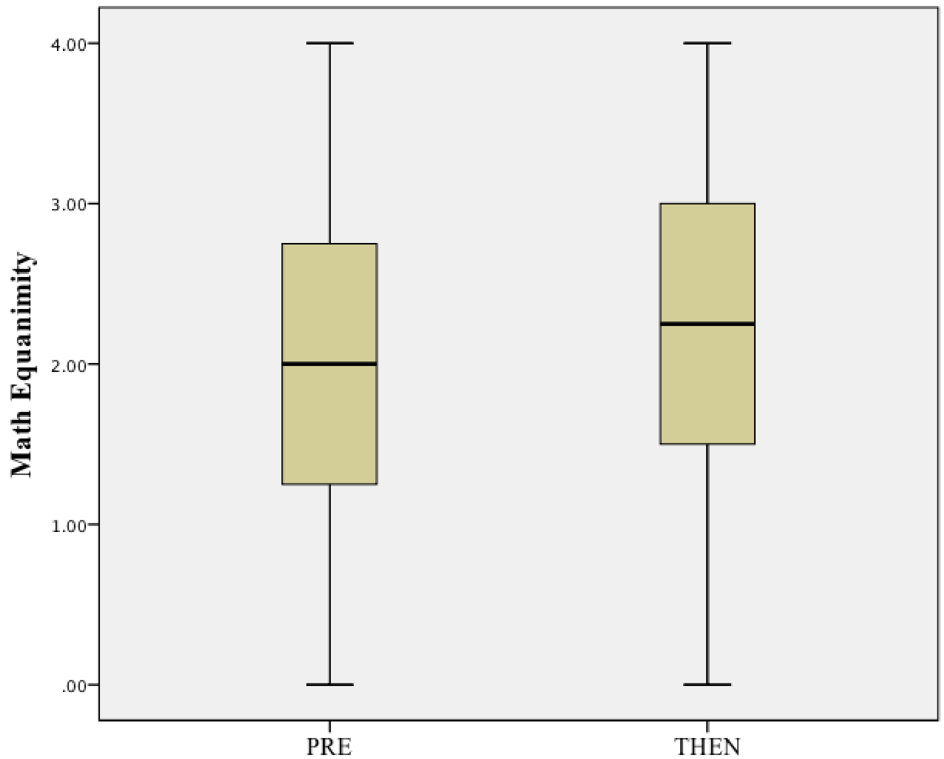


Figure 6. Distribution of pre- and then-survey math equanimity scores

Math Mindset (Pre vs. Then)

As I stated in the results of Research Question 1A, there was not a significant effect of time on students' mindset scores (F (2, 545.43) = 2.82, p > .01); students did not exhibit response shift in their mindset scores from the pre-survey (M = 2.97, SE = .16) to the then-survey (M = 2.98, SE = .16). Figure 7 contains a boxplot of the distribution of math mindset scores on the pre- and then-survey.

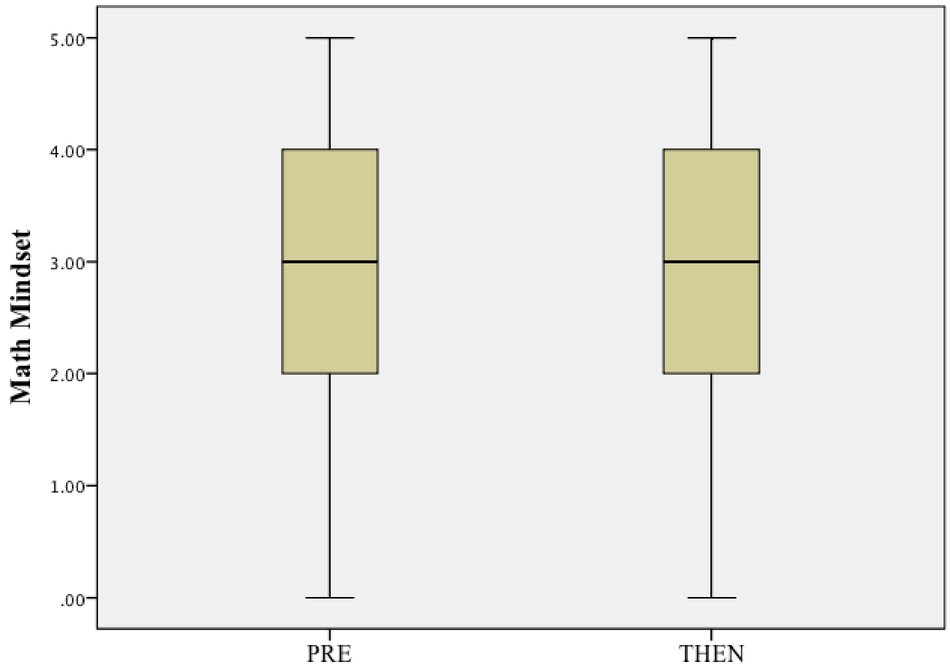


Figure 7. Distribution of pre- and then-survey math mindset scores

Math Self-Efficacy (Pre vs. Then)

As I stated in the results of Research Question 1A, there was not a significant effect of time on students' self-efficacy scores (F (2, 542.73) = 1.75, p > .01)); students did not exhibit response shift in their self-efficacy scores from the pre-survey (M = 2.62, SE = .13) to the then-survey (M = 2.71, SE = .13). Figure 8 contains a boxplot of the distribution of math self-efficacy scores on the pre- and then-survey.

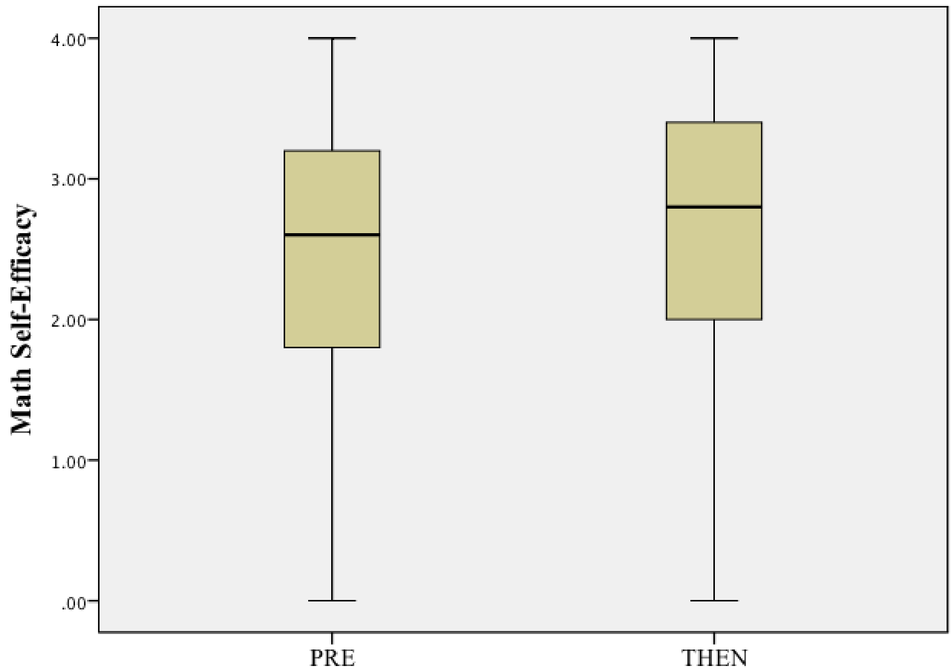


Figure 8. Distribution of pre- and then-survey math self-efficacy scores

Math Belongingness (Pre vs. Then)

As I reported in the results of Research Question 1A, there was not a significant effect of time on students' math belongingness scores (F (2, 542.19) = 2.42, p > .01)); students did not exhibit response shift in their math belongingness scores from the pre-survey (M = 2.58, SE = .15) to the then-survey (M = 2.73, SE = .15). Figure 9 contains a boxplot of the distribution of math belongingness scores on the pre- and then-survey.

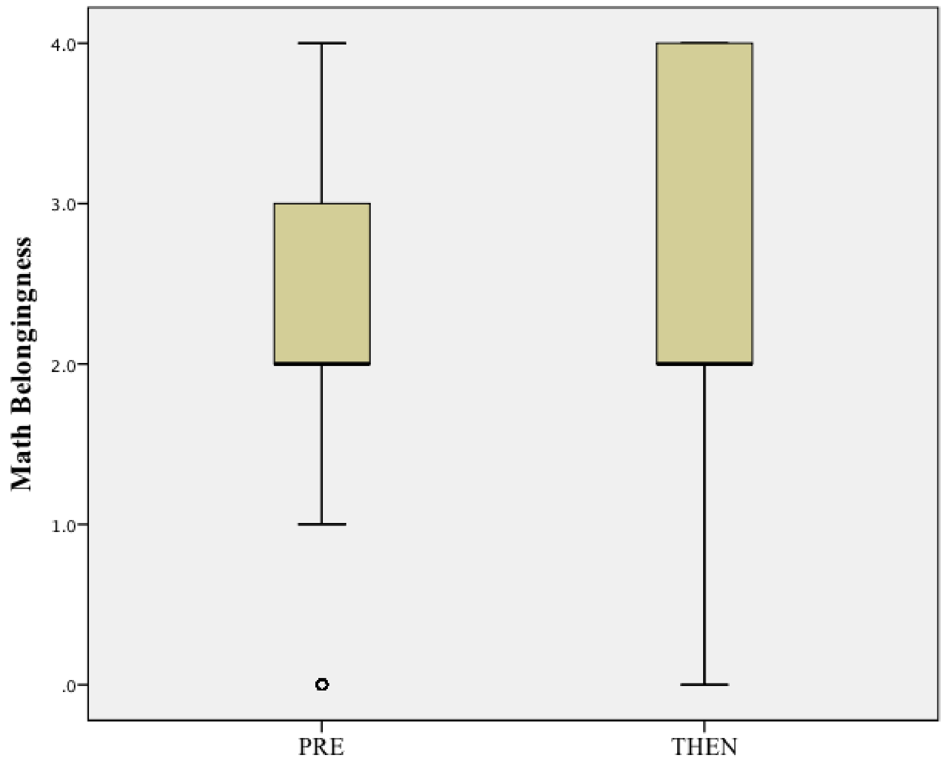


Figure 9. Distribution of pre- and then-survey math belongingness scores

College Belongingness (Pre vs. Then)

As I reported in the results of Research Question 1A, there was not a significant effect of time on students' college belongingness scores (F (2, 536.70) = 3.83, $p > .01$); students did not exhibit response shift in their college belongingness scores from the pre-survey ($M = 2.99$, $SE = .15$) to the then-survey ($M = 2.82$, $SE = .15$). Figure 10 contains a boxplot of the distribution of college belongingness scores on the pre- and then-survey.

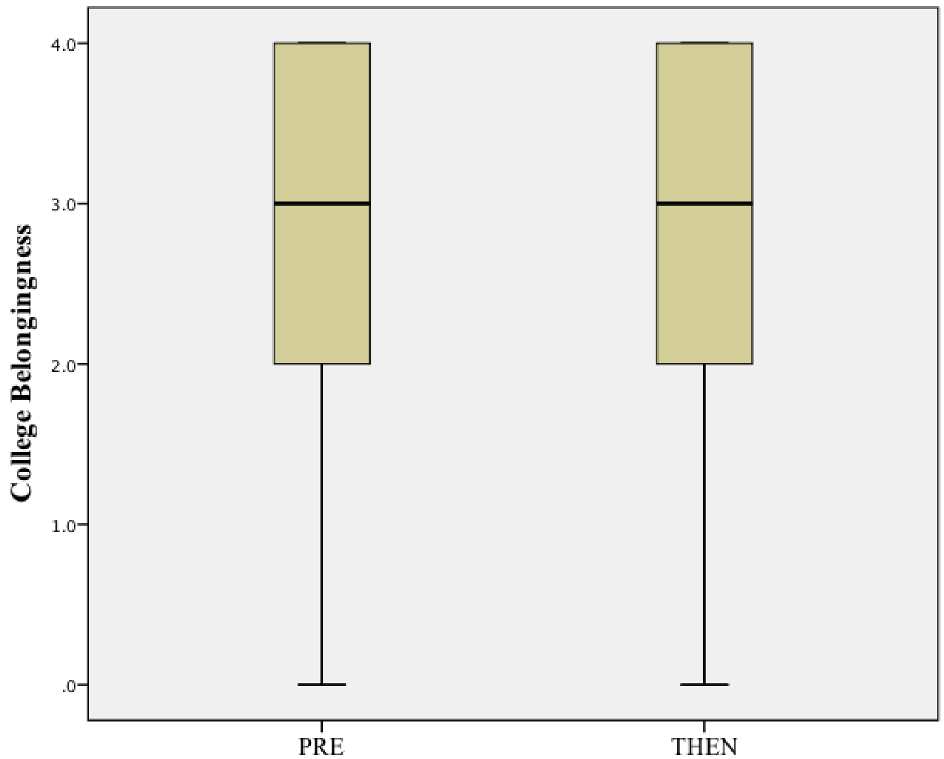


Figure 10. Distribution of pre- and then-survey college belongingness scores

QUESTION 2: OUTCOMES AND CHANGES IN DM-NONCOGNITIVE FACTORS

Do beginning-of-semester to end-of-semester differences in students' math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness predict semester outcomes[40] *(math course grade, math final exam grade, and math course percent attendance)?*

> Research Hypothesis: Beginning-of-semester to end-of-semester differences in students' math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness will predict semester outcomes.

[40] I was unable to use Developmental Assessment of Place Value Understanding scores as a quantitative outcome measure.

To address this research question, I ran three multilevel models, one for each of my dependent variables (math course grade, math final exam grade, and math course percent attendance). I used the change scores in dm-noncognitive factors (post-survey minus pre-survey for each of: math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness) as my independent variables. My models control for pre-survey scores and demographic variables: gender, age, race/ethnicity, enrollment status, job status, number of dependents, native language, maternal education, and first generation college student. I based my significance level on a Bonferroni correction for multiple comparisons and set my alpha level to .017 for all analyses in Research Question 2. I initially treated section as a random effect in the models, but it was not significant for any of the models ($p > .017$) and I did not include section as a random effect in the final models. I used G*Power 3.1.9.2 to conduct a power analysis for Research Question 2 based on power analyses for linear multiple regression. A sample size of 209 subjects is necessary to detect a medium effect size ($f^2 = .13$) with a power of .80 for an alpha[41] of .0125 with 22 predictors; my sample size was adequate to address Research Question 2.

Final Multilevel Models for Research Question 2

The model below was used to test whether or not changes in students' reports of dm-noncognitive factors—math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness—are associated with their class grade. The other two models are parallel to this model, with Course Grade replaced by Final Exam or Attendance. Because I removed the class section variable from the models I described

[41] I originally intended to use four models to address Research Question 2, so my alpha was based on a Bonferroni adjustment for four models.

in the methodology chapter, the resulting model is equivalent to a regression equation and analyses yield the same results as using regression.

Final Equation

$$\text{Course Grade}_i = \beta_0 + \beta_1 * \left(\text{Equanimity}_{\text{Post-pre}}\right)_i + \beta_2 * (\text{Mindset}_{\text{Post-Pre}})_i + \beta_3 * (\text{SelfEfficacy}_{\text{Post-Pre}})_i + \beta_4 * (\text{MathBelonging}_{\text{Post-Pre}})_i + \beta_5 * (\text{CollegeBelonging}_{\text{Post-Pre}})_i + \beta_6 * (\text{School})_i + \beta_7 * (\text{Gender})_i + \beta_8 * (\text{Enrollment})_i + \beta_9 * (\text{Kids})_i + \beta_{10} * (\text{English})_i + \beta_{11} * (\text{FirstGen})_i + \beta_{12} * (\text{Race}_{\text{Black}})_i + \beta_{13} * \left(\text{Race}_{\text{Hispanic}}\right)_i + \beta_{14} * (\text{Race}_{\text{White}})_i + \beta_{15} * (\text{Age})_i + \beta_{16} * (\text{Job})_i + \beta_{17} * (\text{MaternalEd})_i + \beta_{18} * (\text{Equanimity}_{\text{Pre}})_i + \beta_{19} * (\text{Mindset}_{\text{Pre}})_i + \beta_{20} * (\text{SelfEfficacy}_{\text{Pre}})_i + \beta_{21} * (\text{MathBelonging}_{\text{Pre}})_i + \beta_{22} * (\text{CollegeBelonging}_{\text{Pre}})_i + \varepsilon_i$$

In the above equation, Course Grade_i is the class grade for the i^{th} student and β_0 is the predicted mean of class grade.

Summary Results for Research Question 2

Table 6 lists results from the tests of fixed effects for each of the three models used to address Research Question 2. The table includes the significant and non-significant variables of interest (change scores for math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness), as well as the significant control variables. Control variables that were not significant for a model are excluded from the results presented for that particular model, $p > .017$. The sections following Table 6 provide more information about the significant variables in each model.

Table 6

Tests of Fixed Effects for Variables of Interest and Significant Control Variables in Research Question 2 Models

Models	Variables	df_1, df_2	F	p
Math Course Grade				
(*n*=247)	Equanimity change	1, 224.00	5.36	>.017
	Mindset change	1, 224.00	5.19	>.017
	Self-Efficacy change	1, 224.00	9.67	.002
	Math Belonging change	1, 224.00	0.26	>.017
	College Belonging change	1, 224.00	2.13	>.017
	Equanimity pre-survey	1, 224.00	6.63	.011
	Self-Efficacy pre-survey	1, 224.00	19.27	.000
	School	1, 224.00	32.85	.000
	Gender	1, 224.00	15.68	.000
Final Exam Grade (*n*=237)				
	Equanimity change	1, 214.00	2.95	>.017
	Mindset change	1, 214.00	1.84	>.017
	Self-Efficacy change	1, 214.00	12.28	.001
	Math Belonging change	1, 214.00	0.93	>.017
	College Belonging change	1, 214.00	1.24	>.017
	Self-Efficacy pre-survey	1, 214.00	11.68	.001
	Gender	1, 214.00	7.04	.009
Attendance (*n*=243)				
	Equanimity change	1, 220.00	8.46	.004
	Mindset change	1, 220.00	0.44	>.017
	Self-Efficacy change	1, 220.00	1.79	>.017
	Math Belonging change	1, 220.00	0.12	>.017
	College Belonging change	1, 220.00	0.24	>.017
	Equanimity pre-survey	1, 220.00	12.75	.000

Note. This table includes results for both significant and non-significant variables of interest for semester outcome models—change scores for math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness—as well as significant control variables. Significance was set at $p < .017$.

Detailed Results for Question 2

Math Course Grade

There was a significant effect of self-efficacy change scores on students' course grades; for every unit increase in self-efficacy change score, a 3.78 increase in course grade is expected. Table 7 lists the coefficients for the significant continuous variables in the model for Math Course Grade, and Table 8 lists the estimated marginal means of the significant categorical variables in the model for Math Course Grade.

Table 7

Coefficients and Standard Errors of Significant Continuous Variables for Math Course Grade

Variables	Coefficients	*SE*
Self-Efficacy	3.78	1.21
Equanimity pre-survey	-3.13	1.22
Self-Efficacy pre-survey	6.04	1.38

Note. Significance was set at $p < .017$.

Table 8

Estimated Marginal Means and Standard Errors of Significant Categorical Variables for Math Course Grade

Variables	*M*	*SE*
School		
College A	73.66	2.11
College B	84.40	2.34
Gender		
Male	75.16	2.51
Female	82.90	1.95

Note. Significance was set at $p < .017$.

Final Exam Grade

There was a significant effect of self-efficacy change scores on students' final exam grades; for every unit increase in self-efficacy change score, a 4.69 increase in final exam grade is expected. Table 9 lists coefficients for the significant continuous variables in the model for Final Exam Grade, and Table 10 lists the estimated marginal means of the only significant categorical variable in the model for Final Exam Grade, gender.

Table 9

Coefficients and Standard Errors of Significant Continuous Variables for Final Exam

Variables	Coefficients	*SE*
Self-Efficacy change	4.69	1.34
Self-Efficacy pre-survey	5.30	1.55

Note. Significance was set at $p < .017$.

Table 10

Estimated Marginal Means and Standard Errors of Significant Categorical Variable for Final Exam Grade—Gender

Variable	*M*	*SE*
Gender		
Male	77.52	2.80
Female	83.43	2.15

Note. Significance was set at $p < .017$.

Attendance

There was a significant effect of equanimity change scores on students' percent attendance; for every unit increase in equanimity change score, a decrease of 1.87 percentage points in attendance is expected. Table 11 lists coefficients for the significant continuous variables in the model for Attendance. There were no significant categorical variables, $p > .017$.

Table 11

Coefficients and Standard Errors of Significant Continuous Variables for Attendance

Variables	Coefficients	*SE*
Equanimity change	-1.87	0.64
Equanimity pre-survey	-2.66	0.75

Note. Significance was set at $p < .017$.

QUESTION 3: EVIDENCE OF PLACE VALUE CONCEPT TRANSFER

Do students exhibit evidence of their ability to transfer their knowledge to novel place value problems?

Research Hypothesis: Students will exhibit evidence of their ability to transfer their knowledge to novel place value problems.

To shine light on this question, I analyzed the students' online DAPVU responses in terms of 4 facets of place value understanding: Accuracy, Representation, Descriptive Language, and Depth. I assigned students' responses to evidence marker categories associated with these facets. Full descriptions of the markers are in Appendix O. I also considered response rates and time spent on the assessment. There are 5 problem situations on the DAPVU. Out of the 70 students who opened the DAPVU, 45 saw Problem Situation 1, 42 saw Problem Situation 2, 33 saw Problem Situation 3, 28 saw Problem Situation 4, and 28 saw Problem Situation 5. Two students saw the entire DAPVU, but did not respond to any of the questions. A total of 43 students, 26 from College A and 17 from College B, responded to at least 1 problem situation. Out of the

26 College A students who responded, 10 had the same instructor. Out of the 17 College B students who responded, 10 had the same instructor. The rest of the respondents were more evenly distributed across instructors. The demographic backgrounds of DAPVU respondents are in Appendix I.

Students were asked to spend no more than one hour on the assessment. The majority of respondents took much less than one hour. Table 12 compares the amount of time students took on the assessment to how many problem situations for which they provided a response. Students could choose to skip problems; so, a student who recorded responses on two problem situations may have recorded responses only on Problem Situation 1 and Problem Situation 5, for example.

Table 12

Comparison of Amount of Time Students Spent on DAPVU to Number of Problem Situations Attempted

	Number of Attempted Problem Situations				
Total Time	1	2	3	4	5
1-5 min. (*n*=2)	1	1			
5-10 min. (*n*=11)	1	6	1		3
10-15 min. (*n* =3)		1			2
15-20 min. (*n*=7)		1	1	1	4
20-25 min. (*n*=3)		1			2
30-35 min. (*n*=4)			1	1	2
50-55 min. (*n*=1)					1
55-60 min. (*n*=1)					1
90-95 min. (*n*=1)					1
120+ min. (*n*=10)			2	3	5
Totals (*n*=43)	2	10	5	5	21

Detailed Results for Question 3

Problem Situation 1: Pat's Skiing Competition

Forty students responded to PS1. I categorized student responses to PS1 using the Accuracy, Representation, and Depth (of understanding) facets. Thirty-five students provided inaccurate computations and 4 provided accurate computations. One student provided a written response, "add the time together", but did not provide calculations. Nine students represented their responses in terms of digits and units (e.g., 5 minutes, 44 seconds, 21 milliseconds), 13 students used fully symbolic nonstandard representations (e.g., 5:44:21), and 17 used fully symbolic standard representations (e.g., 5:44.21).

For Depth (of understanding), I did not have information to classify 9 students: I could not determine the computation method for seven students and two students used estimation. According to Hannigan (1998), "computational techniques unique to mental computation and estimation are facilitated by a strong understanding of place value" (p. 44). The students who estimated may have a very sophisticated understanding of place value, but it was not possible to decipher whether or not this was the case. The remaining 31 students were placed in the ordinal evidence markers for depth of understanding on PS1, with 0 representing "no evidence that the mixed-grouping place value structure is recognized" and 5 representing "evidence that the mixed-grouping place value structure is recognized and utilized." A student with a score of 5 does not have computational errors. One student evidenced a Depth score of 0, 11 students evidenced a score of 1, 7 students evidenced a score of 2, 6 students evidenced a score of 3, 2 students evidenced a score of 4, and 4 students evidenced a score of 5.

Problem Situation 2: Bobby's Squares

Thirty-eight students responded to PS2. I categorized student understanding on PS2 in terms of their use of descriptive language and their depth of analysis. The ordinal evidence markers for descriptive language use on PS2 range from (0) "no or inaccurate use of place value language" to (4) "accurate and highly specific use of place value language." Twenty-six students evidenced a Descriptive Language score of 0, 3 students evidenced a score of 1, 1 student evidenced a score of 2, 4 students evidenced a score of 3, and 4 students evidenced a score of 4. The ordinal evidence markers for Depth (of analysis) on PS2 range from (0) "provides an irrelevant, incorrect, or uninformative analysis" to (4) "develops an accurate and elaborate analysis of the place value concepts not understood by the child." Twelve students evidenced a depth score of 0, 18 students evidenced a score of 1, 2 students evidenced a score of 2, 4 students evidenced a score of 3, and 2 students evidenced a score of 4.

Problem Situation 3: Chocolate Factory

Thirty students responded to PS3. I used 3 facets of place value understanding to categorize PS3 responses: Accuracy, Representation, and Depth (of understanding). Twenty-six students provided inaccurate computations and no students provided accurate computations. Four students explicitly expressed confusion in written responses, but did not provide calculations. One of these students suggested that more details, a graph, or a picture might help him or her solve the problem. Another student said it would be helpful to see someone else work the problem first. Six students represented the quantity in terms of digits and units (e.g., 3 boxes, 2 packages, 2 singles) and the other 20 students used symbolic representations with digits only (e.g., 322).

I did not have enough information to classify 14 students' depth of understanding: I could not decipher the computation method for 7 students, 3 students merely restated the initial quantity with units (3 cartons, 0 boxes, 1 package, & 2 single chocolates) or restated the example quantity, and 4 students explicitly expressed confusion (same as mentioned above). The remaining 16 students were placed in the ordinal evidence markers for Depth (of understanding) on PS3, ranging from 0 to 4. A student with a score of 0 showed no evidence that he or she recognized or utilized the base-four place value structure and may have applied base-ten strategies in inappropriate situations, indicating significant conceptual errors. A student with a score of 4 showed evidence that he or she recognized and utilized the base-four place value structure and did not make computational errors. Eleven students evidenced a depth score of 0, 5 students evidenced a score of 1, and no students evidenced scores of 2, 3, or 4.

Problem Situation 4: Rugolian Rug Merchant

Twenty-six students responded to PS4. I used evidence markers for the Accuracy, Representation, and Depth (of understanding) facets to categorize student responses on PS4. Nineteen students provided inaccurate computations and 2 provided accurate computations. Five students provided written responses expressing confusion (e.g., "I'm so lost"), but did not provide calculations. Seven students represented their results in the form of digits and units (e.g., 7 blue, 1 red, 1 green, 0 yellow), 4 students used letters only (e.g., BBBBBBBRG), no students used a fully symbolic representation (e.g., 7:1:1:0), and 10 used a different (seemingly nonsensical) representation (e.g., $200).

For depth of understanding, I did not have enough information to classify 16 students: I could not determine the computation method for 7 students, 3 students appeared to believe the prompt asked for which of the 2 rugs was more/less expensive as opposed to the sum of the 2 rugs, and 6 students explicitly expressed confusion (one in addition to the 5 mentioned above). The remaining 10 students were placed in the ordinal evidence markers for depth of understanding on PS4, with 0 representing no evidence that the mixed-grouping place value structure is recognized and 4 representing evidence that the mixed-grouping place value structure is recognized and utilized. A student with a score of 4 uses efficient regrouping strategies, represents the quantity using the fewest number of coins possible, and does not have computational errors. One student evidenced a depth score of 0, 6 students evidenced a depth score of 1, 1 student evidenced a score of 2, 1 student evidenced a score of 3, and 1 student evidenced a score of 4.

Problem Situation 5: Maria's Error Pattern

Twenty-eight students responded to PS5. I categorized student responses to PS5 in terms of accuracy, use of descriptive language, and depth of analysis. Twenty students produced computations that did not involve reproduction of Maria's error pattern. Four students submitted partially accurate reproductions of the error pattern and 2 students accurately reproduced Maria's error pattern. Two students provided written responses (e.g., "I don't understand her thinking at all"), but did not provide calculations. The ordinal evidence markers for descriptive language use on PS5 range from (0) "no or inaccurate use of place value language" to (4) "accurate and highly specific use of place value language." Fourteen students evidenced a descriptive language score of 0, 9 students evidenced a score of 1, 1 student evidenced a score of 2, 3 students evidenced a score of 3, and no students evidenced a score of 4. I did not have enough information to classify 1 student in terms of descriptive language use (the student provided calculations, but no analysis of Maria's thinking).

The ordinal evidence markers for depth of analysis on PS5 are identical to those used in PS2 range from (0) “provides an irrelevant, incorrect, or uninformative analysis” to (4) “develops and accurate and elaborate analysis of the place value concepts not understood by the child.” Ten students evidenced a depth score of 0, 13 students evidenced a score of 1, 2 students evidenced a score of 2, 1 student evidenced a score of 3, and 1 student evidenced a score of 4. I did not have enough information to classify 1 student’s depth of analysis (the student provided calculations, but no analysis of Maria’s thinking).

Evidence for Facets of Place Value Understanding

In this section, the results for Research Question 4 are organized in terms of facets of place value understanding. Tables 13 through 16 contain frequencies for the accuracy, representation, descriptive language, and depth ordinal evidence markers. Note that evidence marker scores for one problem situation may not be directly comparable to scores for another problem situation on the same facet. For example, Depth has six evidence markers for PS1, but only five evidence markers for PS3; the markers indicating the greatest depths on the two problem situations have different anchors. Hence, the tables should be viewed in terms of the spread of scores. Refer to Appendix O for a full description of the evidence makers for each problem situation.

Table 13 contains frequencies for student accuracy on Pat’s Skiing Competition (PS1), Chocolate Factory (PS3), Rugolian Rug Merchant (PS4), and Maria’s Error Pattern (PS5). The majority of students did not provide evidence of accuracy. On PS1 and PS4, only approximately 10% provided accurate responses. On PS5, 8% were accurate and 15% were partially accurate. No students provided accurate responses on PS3.

Table 13

Distributions of Students' Accuracy by Problem Situation

	Problem Situations			
Evidence Markers	PS1	PS3	PS4	PS5
Inaccurate/no reproduction of error pattern	35	26	19	20
Partially accurate	/	/	/	4
Fully accurate	4	0	2	2
Totals	39	26	21	26

Note. A / indicates that the problem situation does not utilize the corresponding evidence marker.

Table 14 contains frequencies for students' choice of representation on Problem Situations 1, 3, and 4. These numbers include only the students who provided quantitative representations. On Pat's Skiing Competition (PS1) and Chocolate Factory (PS3) Problem Situations, approximately 23% represented their responses with digits and units (e.g., 5 minutes, 44 seconds, 21 milliseconds or 3 boxes, 2 packages, 2 singles) and 77% of students represented their responses symbolically (e.g., 5:44.21 or 322). On the Rugolian Problem Situation (PS4), approximately 33% students represented their responses in terms of digits and units (e.g., 7 blue, 1 red, 1 green), 19% used letters only (e.g., BBBBBBBRG), no students represented their responses symbolically (e.g., 7:1:1:0), and 48% used a seemingly nonsensical representation (e.g., $200).

Table 14

Distributions of Students' Representations by Problem Situation

	Problem Situations		
Evidence Markers	PS1	PS3	PS4
Digits and units	9	6	7
Letters only	/	/	4
Symbolic	30	20	0
Other (nonsensical)	/	/	10
Totals	39	26	21

Note. A / indicates that the problem situation does not utilize the corresponding evidence marker.

Table 15 contains frequencies for students' use of descriptive language on Bobby's Squares (PS2) and Maria's Error Pattern (PS5). Very few students used specific place value language. When students used non-specific place value language, it was generally used to indicate behaviors, not concepts. For example on PS5, one student who replicated the error pattern explained, "Maria is carrying the first digit wrong by taking away how ever many she borrowed from the rest of the digits." A student who did not correctly reproduce the error pattern on PS5 wrote, "Maria is borrowing too many from the neighboring numbers." When discussing what Bobby does not understand on PS2, a student who used specific place value language wrote, "I don't think he understands that two isn't actually two it actually means twenty." A student using highly specific place value language on PS2 explained that Bobby doesn't understand "the two is in the tens place and the 6 is in the ones."

Table 15

Distributions of Students' Descriptive Language Use by Problem Situation

	Problem Situations	
Evidence Markers	PS2	PS5
No or inaccurate use of place value language	26	14
Accurate, non-specific place value language used to indicate a behavior	3	9
Accurate, non-specific place value language to indicate a concept	1	1
Accurate, specific place value language	4	3
Accurate, highly specific place value language	4	0
Totals	38	27

Table 16 contains frequencies for students' depth of understanding PS1 (Pat's Skiing Competition), PS3 (Chocolate Factory), and PS4 (Rugolian Rug Merchant) and frequencies for students' depth of analysis on PS2 (Bobby's Squares) and PS5 (Maria's Error Pattern). Note that PS1 has six evidence markers and PS2-PS4 have five evidence markers. For all problem situations, the majority of students did not provide more than partial procedural understanding. As stated above, the evidence markers for depth are not equivalent across problem situations (except PS2 and PS5). A score of a 1 on PS4 may show stronger evidence of depth of understanding than the same score on PS3. For example, one student who received a 1 (likely due to a calculation error) on PS4 described, "I used the table and lined up the same coins in each column from the question then started crossing off coins that could make less coins like 4 yellows made green." Another student received a 1 on PS3 due to a significant conceptual error. He or she represented a full case in base-four as 4444, instead of 10,000. Because I was unable to see the students' work, I was not always able to determine their computation method. If I had access to this information, their scores may have been higher.

Students evidenced greater depth of understanding on PS1 and PS2 and this could, in part, be due to the fact that they were explicitly asked to describe concepts, not just their own problem-solving processes. Other differences between students' depth of analysis or understanding could stem from the type of quantitative representation in each problem situation. The two problems where students evidence the greatest depth both have familiar-systematic quantitative representations. Students exhibited greater spread of depth on PS1 than on PS3 and PS4; the former has a familiar-nonsystematic quantitative representation and the latter two have unfamiliar-systematic and unfamiliar-nonsystematic quantitative representations, respectively.

Table 16

Distributions of Students' Depth of Analysis/Understanding by Problem Situation

	Problem Situations				
Evidence Markers	PS1	PS2	PS3	PS4	PS5
0 ~ No evidence of understanding	1	12	11	1	10
1 ~ Evidence of partial procedural understanding	11	18	5	6	13
2 ~ Evidence of partial conceptual understanding	7	2	0	1	2
3 ~ Evidence of conceptual understanding	6	4	0	1	1
4 ~ Evidence of conceptual understanding	2	2	0	1	1
5 ~ Evidence of conceptual understanding	4	/	/	/	/
Totals	31	38	16	10	27

Note. A / indicates that the problem situation does not utilize the corresponding evidence marker.
Scores from 3 to 5 range from conceptual understanding with a minor conceptual or computational error to conceptual understanding with no errors. A score of 4 on PS2, PS3, PS4, or PS5 indicates no errors.

SEMI-STRUCTURED INTERVIEW RESULTS

Three students participated in one-on-one, audio-recorded, semi-structured interviews: Mia, Casie, and Robert. I provided their demographic information in Chapter

3. The interview template and think-aloud problems are in Appendix P. All three interviews lasted longer than one hour. The interviews were conversational in nature and the interviewees appeared very relaxed and open to share their ideas. Casie and Robert were especially talkative. Casie said she agreed to the interview because, even though she felt “kinda fried” at the end of the semester, she knew she would be bored during the semester break when she’s not able to learn as much. Robert said he agreed to the interview because he did not think I would be able to convince many people to do an interview at the end of the semester. Because the different pieces of the interview template are interrelated, they all answered some of my intended questions prior to my asking. I have broken this section down into interview themes and have provided representative statements from the students. Because the dm-noncognitive factors and outcomes are intertwined, a student’s statement under one theme may also apply to another theme.

Background

All three students described themselves as older students because they had taken several years off, at some point. They all seemed happy about being in classes. When asked about her background, Mia disclosed information related to personal life struggles and described how she had to climb out a very difficult situation. Throughout the interview, she mentioned how religious faith has helped her tremendously by relieving anxiety and steering her life in a more positive direction. She said she didn’t do anything except sleep during her high school math classes. After not taking a math class for 10 years, Mia was placed in Foundations via a placement test. She said she did not initially remember much math, but remembered more as the semester progressed. She felt good to be in college: “It’s about time! I was always in confusion about what I wanted to do until I got into a relationship with God and he told me what I needed to be doing.”

Casie was a straight-A student through 8th grade. Before high school, everything came easily to her, but she wasn’t motivated during high school because “it didn’t seem like it mattered” and she couldn’t get the Advanced Placement classes she desired. She thought she was smart enough, but the message she received from the administration was, “You’re not smart enough.” She said her teachers didn’t care whether or not the students completed assignments and she stopped doing her schoolwork. A new state exam had been implemented and the teachers were just teaching to the test. She wanted to learn the rest of the course material, but the teachers were focused on just material that would be on the exit exam. She questioned, “Is this really school?” She was not interested and, because she thought the teachers didn’t care, it was difficult for her to care. She began homeschooling her sophomore year in high school to care for her ill mother. A teacher would go to her house for one and one-half to two hours each week and then Casie would do her homework alone. She said she only took three years of math in high school, barely passing her classes and not learning anything. She attributed this to apathy, being homeschooled, and thinking she would not need math in the future. Casie tried to go to college after high school, but said she dropped out within the first two weeks because she was not ready and due to circumstances outside of her control. She said she was approaching her ten-year anniversary of graduating from high school and, about her return to school, she said, “I’m finally back.” She remembered a lot from high school after “a little jogging,” but “nothing from math…because there was just nothing there to remember.”

Robert had moved from another state and was taking classes to transfer back to his home state for an associate degree, but he said he avoided math for as long as he could. He had taken one math class at another Texas community college and one math class at College A prior to taking Foundations at College A. Speaking of his math class at the first Texas college he attended, he said: "You were afraid to because you didn't want to be called out in class and feel like an idiot…We didn't get it the way he thought we should get it; So, his feelings were being projected that we weren't good enough to do it, that we should be doing this more, that we should be working harder. I understand you have to work hard to succeed in the class, but if you're starting out on the fourth step and not doing the first two well enough you are going to stumble on the third one."

Foundations Instructors

Mia, Casie, and Robert claimed to have learned a lot and volunteered much praise about their instructors. Mia and Casie had the same Foundations instructor, but they were in different sections. Mia said, "He was funny and nice, explained everything, and had a lot of jokes—he was more interesting than most other math teachers." During the think-aloud portion of the interview, Mia expressed that the DAPVU and think-aloud problems were foreign to her and said, "A lot of stuff in my math class I didn't get it until he showed me and then I got it." Mia claimed that she learned a lot because of the way the instructor interacted with the students and made the content relevant to their lives. Casie said he was the best teacher she had ever encountered, in general, not just in math. Casie talked about how he was very patient and clear when explaining problems on the board, and how he repeated things when the students needed him to do so.

Casie contrasted her Foundations instructor with a prior high school English teacher who would call absent students during class to harass them and make fun of them in front of the other students. Before facing that teacher, English was her favorite subject. When she requested help after falling behind in a high school math class due to excused absences, her teacher's response was, "Well, you should have been here." Her mental response to this was, "Well, if they don't care, I don't care." In her Foundations course, Casie felt like it was important to do and understand the math, and part of this was because she thought the instructor really cared and wanted everyone to get As.

Robert had the same instructor in Foundations as his math class at College A that immediately preceded Foundations. He enjoyed the instructor's teaching style. He said, "In my experience, if a math teacher's good, they're able to express ideas in a way students can understand it. If they're just doing it in a very rigid way, you're going to have people that don't get it and they are going to struggle with it and have associated bad feelings towards it."

Math Interest, Utility of Foundations, and Future Plans

Mia's view of math changed significantly during Foundations, saying her ability to solve problems and be successful in the course made her like math. She talked about how seeing real world examples drove home the point that "we need math to live in the world." She explained: "It wasn't just about the math. He made it about normal life stuff…and gave examples about our lives. The work we did had a lot of examples in statistics and the studies they've done; so it was personal. He made it more real…and showed us how we would use it in our lives or in other classes in college." She said the lessons utilizing graphing and spreadsheets for algebra would be immediately useful in her current job. Mia intended to take the next math course in the sequence the following semester, but could not remember the course name. She said it was not algebra; so, I believe she was enrolled in Contemporary Math. Mia planned to become a marriage and family counselor.

Casie planned to become an epidemiologist (researcher or professor) and knew math will be useful for her career. She said it is urgent for her to understand a lot of math to be successful—"It's like, I need to learn it. I need to learn this now." I asked her if she felt like she focused more on the grades or the learning. She said, "It started off as the grade. I wanted to have the best grades in the class so I could have good transfer scores whenever I go on, but then I was just like, 'the only way I'm going to get good grades is if I learn it well enough to keep doing it'…It was kinda fun to learn it anyway. So, it shifted from need to have perfect grades to need to learn this well."

Casie also believed Foundations content is something she can use in the real world: "He made it relatable. I was really bad at geometry, but he did give us some geometry problems, like finding the area of stuff…[He would say], 'You're finding the area of your backyard because you're renovating,' and it's like, well, maybe I'm going to need to do that or like hang curtains or put up wallpaper or something. You find the area so you know, 'I need to hang this much.' I would need to know that cause, otherwise, I would just throw fabric up and hope for the best." She also appreciated that the instructor taught them memory devices, such as FOIL[42] and PEMDAS[43], to use in Foundations and future courses. Casie described her general math interest transformation as follows: "I wasn't really excited to take a math class in the first place. I was like, 'Ugh, I hate math,' but now, I'm just like, 'Math is my subject!' It was just 180." At one point in the interview, she exclaimed, "I just have so many good things to say about this class!" At

[42] FOIL is a mnemonic device for multiplication of two binomials; it stands for first, outer, inner, last.
[43] PEMDAS is mnemonic device for order of operations; it stands for parentheses, exponents, multiplication, division, addition, subtraction.

the end of the interview, I asked if there was anything else she wanted to say about the class, she said, “I hope everyone takes it, every freshman…seniors, people higher in college, they just go back and take it because it was great.” She had enrolled in the fast math track for the subsequent semester—she intended to take Intermediate Algebra for the first half of the semester and College Algebra for the second half of the semester.

Robert said he had always struggled with math and said he tends to dislike a subject when he finds it difficult: “For me, I like math when I understand it. When I struggle with it, it is the worst thing in the world. I’d rather sit and write 30 page essays than do math.” Grades and comprehension were both important to Robert—“The goal is to get a good grade in the class, but on a personal level, it’s important to understand what you’re learning because, if you don’t understand it, you’re wasting your time. You are paying for enlightenment almost.” While he appreciated its utility and desired to understand it, he did not have much interest in math outside of his coursework; he predominately viewed math as a means to an end. According to Robert, math classes are usually “the most dry” out of all the subjects.

Foundations was the second math course Robert took with his Foundations instructor and he described her as an excellent teacher. Robert claimed the Foundations course was a totally different experience. Foundations was more “more creative than a traditional math class” and there were more peer-to-peer and student-to-instructor interactions; so, students were not left to work in solitude. He explained how the instructor probed students to understand their thinking during Foundations: “When we came up with an answer that didn’t make sense or was the right answer and she just didn’t know how we came to it, she would make us explain it and, in that way, she would have a better idea of how our thought process was with it and what different ways we could use to actually solve it."

Robert contrasted Foundations with traditional classes, where "three plus one is four and that's the only way you can do it." He claimed, "In the new class, the Foundations class, you could do it that way or you could do two plus two equals four. So it gave you a different way of looking at it and a different way of relating it to real life…It wasn't just memorizing it and just the answers; it was a different way of looking at it." He described how problems using miles per gallon and miles per hour were relevant to his life and how he appreciated being able to come up with solutions, without having to rely on technology. He preferred Foundations to traditional courses "because the old way is more about memorization, memorization, memorization and not everybody is geared for that." He said this different approach helps students who struggle with math because it makes the math "more relatable and easier to understand."

Robert's comments about the course were not all positive. He was not fond of the way preview assignments were presented in MyMathLab. He thought the preview assignments did not provide sufficient information to help students understand how to work problems and students were able to game the system. He said, "You could sneak by if you knew how to work MyMathLab. You can do it and still get the credit and not understand it." He learned more during class when he was able to consult with the instructor. The instructor would explain the concepts, show him how to work problems, and provide multiple perspectives. He was a little disappointed that the class was rushed near the end of the semester because some difficult topics, such as geometry, were only covered superficially due to lack of time.

Robert was very nervous about going into a statistics course the following semester. He said Foundations provided “just a scratch on the surface” of statistics, but the course helped him learn how to use a scientific calculator and he would need this knowledge in a statistics course. Foundations helped him by exposing him to multiple ways to work problems, and learning about different methods reduced his anxiety about statistics. Upon completion of Statistics at College A and two internships in his home state, Robert would receive his associate degree in human services. He described his future counseling role as a way to “help people when they are frustrated and angry about things.” He said he was also “working on bachelor’s [degree] stuff to kill time.”

Student Success Course

All three students took a student success course. Mia and Casie took the New Mathways Project Frameworks student success course the same semester as Foundations and Robert took a different student success course prior to Foundations in a different state. Robert did not discuss any links between Foundations and his student success course. Mia and Casie talked about overlaps between Foundations and Frameworks, both mentioning the video, *How We Learn—Synapses and Pathways.* The Foundations curriculum suggests making connections between the video and the Brain Power lesson about brain neuroplasticity, math mindset, deliberate practice[44], and post-failure successes from persistence.

Mia said it was a video “on the mind and how we remember stuff,” and Casie said, “It had to do with how you learn and how you remember something.” Casie watched the video in Frameworks, Foundations, and her psychology class. According to

[44] Deliberate practice differs from repetitive practice. According to Bryk et al. (2013), “[d]eliberate practice eschews rote repetition for carefully sequenced problems developed to guide deeper understanding of core concepts” (p. 13).

Casie's description of the video, people moved on pulley systems back and forth across a gorge to build a suspension bridge, and each time they went across the gorge, they added to the bridge. At first, it was not very strong—"it was just more rope"—but they were eventually able to walk across the bridge with ease. She compared it to synapses "firing back and forth [to] get a path" and said that it's not just that you are repeating something, but "if you keep doing it, it gets easier and easier and easier." Watching the video three times made it "stick better": "The first time I saw it, I didn't remember it. Then I saw it again, and I was like, 'Oh, we saw that earlier.' It started to stick after that."

Math Equanimity

When I asked about changes in anxiety over the course of the semester, Mia claimed to have anxiety that is not specific to math. The only anxiety that was specific to the course was related to getting good grades on tests and homework. She was nervous on the tests because she had fallen a little behind in the course. She attributed changes in her general anxiety to God's help. About twenty minutes into the interview, I asked, "If I gave you a math problem right now, how would it make you feel?" She responded, "A little nervous because I haven't done any math for a little bit." (It had been five days since she took her final exam, but she had been doing it daily during the semester.) When I probed into why it would make her nervous, she said, "Because I don't know what it is. I haven't prepared." I asked how she would feel if she did the problem, but didn't do well on it. She said, "right now, I would probably be okay." She agreed to try some problems.

Casie was "afraid to go to college after all the horror stories in high school," and she tried to accept that it was going to be an unpleasant experience. At first she was nervous signing up for her math class, worrying that she would have an awful instructor. She wasn't excited to take math, but some of friends told her the instructor was a good teacher and this relieved some of her anxiety. When I asked how she feels when she sees a math problem now, she responded: “The initial reaction is still a little math panic where ‘It’s just numbers, ugh, too many numbers.’ Then it’s like, ‘I can work this out probably. Might as well try’…Initial shock and ‘Okay, I can do it.’ Then I’ll start to work on it and then get frustrated with it and then it’s back up to, ‘Ugh, I can’t do this.’ It has to rollercoaster all the way through.” She explained she prefers to skip problems that make her anxious, and when she returns to them, the anxiety is lessened. She used this technique for a problem on the final exam; the problem had been especially distressing because she encountered it early on, but she was able to complete the problem with less anxiety after she completed the other problems.

Grades, in all subject areas, were a source of anxiety for Casie. She once received a C on a history test, and her first reaction was that it might as well have been an F. Her friends had told her, “You don’t have to get 100% on everything.” She had to force herself to calm down midway through the semester because she was getting so upset about grades. I asked Casie how she would feel if she received a B or a C on an assignment or test in Foundations. She said she would want to go back and relearn the material. The failure would make her more anxious the next time similar problems and she would be upset if she repeated her mistakes, because this would indicate there was something she truly didn’t understand. Casie claimed to have much test anxiety, saying how worried she was about doing well on the final. She said she would be more anxious solving a math test problem than a math problem in general because she would be unable to receive the assistance on a test that she could receive on an assignment. However, by the end of the semester, she was less anxious on math tests than other subjects “because there’s going to be a clear-cut answer.” Describing how her math equanimity increased over the semester, she said, “I went from not nervous at all about the other tests because [non-math tests were] going to be easy to not nervous about math because it was easy.”

Robert was somewhat overwhelmed by the workload: “If you add up the preview assignments, the assignments, the assignments you do after the course, it adds up to a lot of time and that doesn’t even include the stuff you are going over in class…If you have other stuff you are working on, it can be a drag on you academically. So there’s that anxiety that maybe you’re not getting it. Why isn’t this clicking in your head? And that can lead to [anxiety].” He was less anxious when he was able to work at home and take breaks than when he was taking math tests. When taking math tests, he was concerned that he may forget formulas or how to apply them.

Robert noted he became anxious a few times when other Foundations students laughed at him for not being able to explain some concepts, even though he had arrived at a correct solution. He said, “There’s also some of the social anxiety with getting up in front of people and getting something wrong and feeling stupid…Other students found it amusing to listen to me squirm about how I came to the solution.” He added that the Foundations environment was much more welcoming than his previous math classes at another Texas College. In his previous math class, he felt like he and the other students were “berated” by his previous instructor when they were unable to provide quick recall of math facts, and he said the instructor interpreted the students’ lack of expedient recall as a general inability to do math. This made him “feel like [he] wasn’t very good” and “snowballed” his feelings of anxiety. In contrast, he said students in Foundations did not have to be as worried about “shame” and “embarrassment” when they made verbal contributions during class.

Robert was very concerned about taking a statistics course during the subsequent semester, saying he "hates statistics," because he would have a new instructor and it would be "new territory." He did not feel comfortable taking statistics because he was too unfamiliar with it to know what to expect. Some students told him statistics is an easy course and others said, "It's living hell." Signing up for statistics was different, and more intimidating, than signing up for a math class because the content is very different and there are "a lot of different formulas to punch into a calculator." He said, "If I know what to expect, I can prepare for it. It's like jumping off a ship into the water without knowing what the water temperature is…I don't like change."

Math Mindset

I asked Mia, "Do you think you're a math person? Not a math person? Is there such a thing?" Her response was: "I'm kinda on the fence on that, the whole idea about everybody being able to [do math]. I know that everybody's able to learn the same things, but I do think some people are more receptive to different things than others. I never thought I was a math person, but I did well." She said that she is probably not a math person because it is not something that interests her primarily. Mia received gratification from being successful in the course and understanding the material: "To get a good grade, you have to understand [the math]. It all ties in together. You get satisfaction and then you feel good when you get good grades."

Casie adopted a growth mindset during the semester: "I used to think that I'm not a math person, so I'm never gonna learn it…If you think like that you're not gonna learn it, but there's no such thing as a math brain or non-math brain." Her Psychology, Frameworks, and Foundations instructors all included discussions and videos about the math theory of intelligence. She did not buy in to the claims immediately: "They mentioned it earlier on in Foundations and I was starting to think, ‘well, maybe,’ but I still wasn’t completely there yet…That was in my psychology class, too. Math brain, not math brain, English brain, not English brain. But, once it was in Frameworks and I saw I was doing well in math, I was like, ‘okay, well maybe’…It’s whatever you want it to be…It made me feel better than to think I was always going to be bad at math, especially going into a science career. I can’t go into a science career being bad at math or *believing* I'm bad at math.”

Multiple times during the interview, Robert attributed my success (getting a graduate degree and being able to teach math and train teachers at “good universities”) to me being “smart” and “good at math”. Following one of these comments, I asked if believes there is such a thing as a math person and if he considered himself a math person. He response was: “There are definitely people that are more predispositioned for math and science. Am I one of them? No, I’ve never been very confident about my math abilities. It’s always been my lowest scoring subject all the way through middle school.”

When I inquired if Robert thought people can change their math ability by working harder or if math ability is just something that can’t be changed, he responded: “It’s practice…The way I look at math is: if you think about it and you can ponder it in your brain and eventually that little light switch in your head will click on and you will get it, and you’ll be like, ‘Oh, that’s what they mean. This is easy.’ Once you get the concept down, but it’s getting to understand the concept that’s the problem…Math is very black and white. It’s very rigid. There’s a certain way math is. There’s no wiggle room. It’s an exact thing and you have to play math’s rules. And if you don't, math will leave you in the dust as far as grades go…Because I don't know maybe the basics of it, it kinda throws me off a little bit.” I asked if he felt like he was any more of a math person over the course of the semester or if his beliefs about himself were fairly constant throughout, and he said, "I am still not a math person. I can honestly say that. I had to work at it through the whole semester." He said he "was always trying because [math] is not something that comes naturally to [him]."

Math Self-Efficacy

Mia had mastery experiences in Foundations: she received a mid-range A on the final exam and an A in the course. Mia disliked most mathematics prior to Foundations because she didn't have faith in her ability. She now attributes her math failures to lack of effort ("I got irresponsible and fell behind") and her math successes to persistence ("I didn't give up and kept going…I knew if I did my part, I knew I could succeed"). She enjoyed algebra because she felt capable in algebra. She explained: "I think that's the way a lot of people will like something. If they tend to be good at it, it's a nice reinforcement." Her general dislike of math changed after taking Foundations—"I like it now because I know I can do it." She attributed this change mostly to her renewed personal outlook on life. She also felt that doing her homework, putting in effort, and deciding to not give up made her like math more.

Casie also received an A in Foundations. Casie contrasted her success in the course with her lack of success in high school: "I started off with probably a D average from high school math, and I got [higher than 100] on the final. So, I did really well. I did very well. I was so happy. He actually showed us how to figure out our scores and I had [greater than 100 percent] in the class, so I just was. I just was happy. I really learned this semester." Talking about being on the fast track for the subsequent semester, she said, "I did really well in Foundations; so I figured I'd be ready…I learn pretty quickly, but I have to have good teachers, too." She appreciated the positive feedback from the instructor: “Having the teacher really believe in you and especially to tell you, ‘you’re doing well…you’re really shining this semester,’ it makes you think, ‘Oh, I’m doing well.’” Her instructor encouraged persistence and facilitated enactive attainments: When students were incorrect, he would wait until they corrected themselves. Casie said this “was better because [*she*] figured out how to work the problems.” She added, “He was always so happy when we got something right, and he was never upset when we got something wrong.” Casie was self-efficacious in algebra, saying she was the algebra helper when she worked in groups. She believed her classmates also felt quite proud of their accomplishments and they felt comfortable making statements such as, “I said the right answer in class.”

Casie’s efficacy was threatened mostly on tests. When talking about her anxiety related to the final exam, she said, “I was like, ‘I’m gonna fail it.’ I’ve never failed a test, ever, but I still sit there, ‘I’m gonna fail. I’m gonna fail.’” Her efficacy is also threatened when she takes more time on a task than she considers appropriate: The DAPVU was designed to take approximately one hour. Casie said she took “an embarrassing” amount of time” on the DAPVU Chocolate Factory problem situation (PS3), but she had only taken a little over 10 minutes on the entire assessment and only 1 minute, 43 seconds on PS3. I asked her what she meant by this, and she said, “I’m really hard on myself when it comes to doing things, and if it takes me longer than what I think is the right amount of time, I think it’s embarrassing, and it probably isn’t.” Casie’s geometry self-efficacy was lower than other content areas, but this improved somewhat in the Foundations course because the problems related to real world examples and the instructor allowed the students to engage in productive struggle.

Casie entered Foundations believing she was not good at math: “I was just resigned to, ‘let me pass with a C or something,’ because I was so bad at math before.” Casie attributed her successes to her own abilities and good instruction and her failures to poor instruction and lack of motivation. She believed the poor math instruction she received prior to Foundations led her to judge herself as “horrible at math,” but she went on to say, “If I was bad at math, I definitely wouldn’t have done as well as I did.” She described how her self-efficacy decreased in other subjects and increased in math. At the end of the interview, when talking about why she thought everyone (even people who were at higher levels) should take the course, she said, “I have never been better at math than I am now.”

Robert initially earned a high B in the Foundations course. He had been unable to complete some of the online work due to a health issue, and the instructor allowed him to go back and complete some problems. He ultimately received an A in the course. Robert felt rewarded by achieving a high grade in Foundations after putting in a lot of effort and learning. More than once, Robert mentioned how he has to work very hard in math classes; he attributes his math successes to effort and his failures to “not knowing the fundamentals enough.” Describing the “fundamentals” of math, he said, “Every different kind of equation is following rules and, if you aren’t 100% on those rules, is where I tend to find I’m not sure. If I’m 100% sure, then I’m pretty sure I did the problem right. But if there’s any doubt in my mind about what the rules are, then I’m not so confident.”

I conducted the interview with Robert at his college, and it was very quiet because there were very few people around. When I asked Robert how successful he felt he could be if I gave him a math problem, he said the environment and the specifics of the problem would affect his performance—“It’s a school environment. It’s very quiet, which is important to me studying. I kinda associate the college with learning and having to do work, but…I think it would depend more on the math problem itself." He feels more self-assured solving problem types he has encountered multiple times, especially "basics" like addition, subtraction, and multiplication. He said, "the more I've been exposed to it, the more confident I am understanding what's going on and answering the question." He described his correctness on the online DAPVU as "50-50 maybe," saying that he didn't think he was successful on the Pat's Skiing Competition problem situation.

I asked Robert what he had imagined his course grade would be prior to knowing his final grade, and he said, "Horrible. I always think I did horrible…I was hoping for a C…I was just hoping I passed." Before receiving his final exam grade, he thought he had done "okay," maybe receiving a C or B, but he thought it was "nothing to throw [his] hat in the air about." He was surprised by his final grade, but said, "that's my whole philosophy in math, that I think I'm going to do horrible. I'm always going to do horrible. I think I'm going to fail and no one listens to me anymore. Like chicken little. The sky is falling. It's how you feel about it I think because I'm not confident in it and I don't have that confidence. But if you're confident, then it shows, you're like, 'I got an A in this. No problem.' Makes you work harder."

Math and College Belongingness

As I mentioned in the background section, all three interviewees were pleased to be back in college, and they felt a sense of connection with the learning environment. Mia, Casie, and Robert were older than many of their peers. Casie said, "They're all really young. I wouldn't tell anyone my age unless they specifically asked." Robert said some of the students had assumed he was the instructor. Robert pointed out that age differences meant he and the younger students had different perspectives on life. Age seemed to play a minimal role in Mia and Casie's sense of math belonging, and a larger role in Robert's, but group work during class helped them feel more like accepted members of the math community. In Mia's Foundations class, the students decided who was in each group. Mia said she isn't "much of a group person" because she is sometimes hesitant talking to people even though she gets along with them. She liked the people in her group and said group work was "good", but she is "still working on opening up to people." She thought her group members contributed equally without a group leader and everyone's ideas were welcome, even though some were a little quieter than others.

Casie said working in groups was usually optional, but her instructor strongly encouraged it and would sometimes put the students in groups by table so students could work in pairs. Except for when she "wasn't feeling very social," she chose to work in a group with male and female students sitting near her. A group of women she worked with during class regularly worked on problems at a café, but she was unable to join due to her babysitting obligations. She enjoyed working in the groups, especially when she was able to help her group members. They would switch roles based on who had better comprehension of the material. When they did algebra she said she was the helper. Other times, she was able to ask questions and felt comfortable telling the other students when she was "completely lost". Her perception was that all group members felt like their ideas were appreciated and they enjoyed helping the other students. They would relay things such as, "I helped this person and they helped me."

Casie had not experienced much group work in previous math classes. She said most teachers get upset when you are talking during class, but her Foundations instructor would praise students for their contributions when they explained their reasoning behind problem solutions. Casie seemed to have developed a sense of belongingness in

Frameworks and Foundations because everyone was working together to achieve success. She said, “Foundations and Frameworks were definitely not traditional, not as college-y as I expected them to be, where you’ve got this great big lecture hall where you be quiet and listen. It was ‘everyone work together and do well.’” The group work experiences helped assuage her feelings of social awkwardness and isolation: “When it’s a whole bunch of people, I just want to sit there. I don’t want to talk to anyone. In class, with all of us working in groups, working together to figure stuff out, it helped alleviate that a little bit.” The entire class, including the instructor, were all working together to achieve a common goal—Casie said the instructor sometimes claimed to not have the answers to problems worked in advance and the class would solve the problems together. She went on to explain how these experiences can help her in other classes and her career, saying, “I won’t be so afraid to ask the person sitting next to me, ‘What is this and how do I do it?’ That will help out career-wise too…I’m going to be like, ‘I’m stuck on this. What do I do?’ instead of just trying to hide it in the corner.” I asked Casie if she felt a sense of community at the college, in general, similar to the sense of community she felt in Foundations. She told me how she did not feel the same way in her history class because it was lecture-based and the students did not interact much. I, mistakenly, did not probe further to determine if she felt an overall sense of belonging beyond the classroom level. Based on her discussion about her excitement to be back in school and considering one day pursuing a career in academia, I believe she felt a strong sense of college belonging.

Robert did not feel strong ties to the college itself. He explained that he had been a student at four different colleges and he saw college as “a means to getting to where [he] wants to go.” As I mentioned in the self-efficacy section, merely being in a school environment primed him to learn. The goal to learn (or even just get a degree) was a connection with other students that contributed to his sense of college belonging. He generally did not feel like a mathematical outsider in his Foundations course, but his age made him sometimes feel a little "out of place with the other students." He said, "I don't feel like an outsider. I just feel old from the fact that most of the kids coming in, the freshmen, they went to high school together, and they look at me, as I told you, they think I'm the instructor of the class." The instructor needed to leave the classroom during one of the tests, and he warned the students that he had "people in [the class] watching" to make sure they did not cheat—three students turned to look at Robert. He thought his age and his regular stays after class to receive math help contributed to this reaction.

Robert had been quite nervous signing up for Foundations. Students were hesitant to raise their hands during his math class at another Texas college because students who provided incorrect answers were shamed, embarrassed, and made to look "stupid" in front of their classmates. In his previous math class at College A, the students worked mostly on computers, and, when he walked in the door to the Foundations class and "saw that there were no computers to hide behind," he worried about talking to people and "exposing [himself] to being wrong." Unlike his previous experiences, there was a sense of community in Robert's Foundations class, and he was "definitely more comfortable with the environment". He said it was not as lecture-based; "it was more of an open discussion." As the semester progressed, he felt "more comfortable with admitting [when he was] wrong." Even when his answers were incorrect, he "didn't feel like an idiot."

Robert said they worked in groups approximately half of the time during Foundations. He enjoyed "engaging with the other students" when he was in a "group with someone who did work, did the assignment, [and] wasn't just kinda winging it off [his] answers." When he was especially confused, he found the student-to-instructor interactions more useful—he was “not afraid to ask the instructor for help,” and “the instructor could take the time out and really emphasize the point of that lesson.” He conjectured that most of his classmates had similar reactions to the classroom interactions, though he said some might not have been vocal because they were younger students who were worried about being embarrassed in front of their peers. He thought the younger students’ social anxiety impacted their desire to share even when their answers were correct. Others were vocal because “they knew what they were talking about” or they were not concerned about how they were viewed by others. Some did not regularly participate because they were “not 100% committed to the class” and would make up work on their own, using MyMathLab. He said the majority of students were working together “to try to figure out what was going on.”

Math Persistence and Confronting Failure

Mia said she got frustrated on y-intercept graphing problems because she had missed a lot of days and didn’t understand the concepts until the instructor went over them with her. She did not go back and work through the problems on which she had been unsuccessful, but she worked similar problems to study for the final. Prior to Foundations, Mia had been unsuccessful in mathematical pursuits, but she believed her failures were due to lack of effort. She thought she could be successful if she tried, but this was the first time she really persisted in math, and she attributed this new desire to persist to her own personally changed outlook on life. During the think-aloud portion of the interview, she tried to solve the same Chocolate Factory problem situation that was on the online DAPVU, but she did not arrive at a solution. I asked if she would be interested in working on it more, not for me, but for her. She said, “Maybe one day if I had some time and it really got to me that I didn’t know how to do it.”

Casie admitted to being apathetic about high school math and not persisting, but she persisted in elementary and junior high. Casie did not feel like she ever gave up during Foundations. I asked Casie how receiving a B or a C on a homework assignment or test in Foundations would impact her. She said, "It would have started off as an end of the world thing," but she would eventually reason that "it's not so bad" if she could talk to the teacher to relearn the material. Her goal would be to not "mess up the next time" because a subsequent failure would indicate lingering misconceptions. She said, "My giving up would be like, 'Hey, I need help,' but that's pretty much it. I worked really hard to work through the problems because I wanted to understand the math for once…it felt like it mattered." Casie productively persisted when she was unable to ask for help on her final exam—"I was just like, 'I have no idea how to do this.' So, I started to panic a little, but then I was like, 'let me just come back to it later.' I finished the rest of it and came back to it, and I was like, 'it's fine.'" She initially had panic on the DAPVU Pat's Skiing Competition problem. She thought, "I'm going to have to add times, and I'm going to mess it up," but she said she "calmed down enough to make an attempt." When I asked if there was anything she wanted to say about the problem situations in general, she said, "The problems were really weird, but they make you think…about why things work a certain way and why they don't work a certain way and I really want to figure that out." She even asked for a copy of PS5 (Maria's Subtraction) because she said it would bother her if she were not able to study it further.

Robert said that he gave up a little near the end of the semester because he knew he was going to receive an A in the course, and he was disinterested in some of the subject matter, like preparing taxes: "I have no interest in that; that's why I have an accountant. So I had no interest in learning it." I asked Robert how he feels when he encounters a problem similar to a problem on which he was previously unsuccessful. He said, “That depends. If I don’t understand the entire concept behind it and I’m just trying to wing it with how I think it is or if I know how to do it and I'm just missing one or two steps. So, I would feel much better if I knew the process and how to do it and maybe I just missed a step and I got confused with another. Whereas, if I didn’t understand it, I would be very [apprehensive] about it.” When I probed about how long he is willing to persist, he said it depends on the depth of the material and how quickly it is being introduced. He gets more frustrated when multiple difficult concepts are being introduced back to back, but his vexations do not necessarily make him quit: “I don’t know if I would say I gave up completely…I’ll take breaks if I don’t get it. I’ll work at it for a while and take breaks and come back to it. Usually if I take a long enough break or if I am staggering the hour into smaller chunks, I tend to find that I get the concepts a little easier. At least I don’t get as frustrated…For the most part, I will keep going until I get it.”

Robert said he had this same mentality about persistence from beginning to end of semester, explaining, “I’m an older student. I want to get stuff. I’m paying the money for it. I want to get it.” Some students fell behind in the course and dropped out, but he thought most of the students were really trying. Even when he struggled, he never considered dropping Foundations because he would “have to do it again next semester and then [he would] be behind.” He added, “I will say what made me feel better sometimes was, if I did fail, if there was a chance I did not pass this class, then at least I would have a foundation to build off of for the next semester. So it would be easier going into it again.”

Robert was in an airport when he began the online DAPVU. He said it started "okay," but he "got to a point where [his] head started hurting and [he] was like, 'I'm done with this.'" He explained that his bad mood and being tired "affected his ability to think out the problems." He stopped at the Rugolia problem. In the airport, "there was too much stimulus…and it was hard to concentrate." When he returned from his trip and started the assessment again "in a quiet house," he "thought it was much clearer than trying to do it in the airport on [his] phone."

Interviewees' Surveys Compared to Interview Responses

Mia said the instructor explained the format of the post-then-survey and it was easy to understand. Robert also felt comfortable with the format, saying, "It made sense because you want to get how they feel when they first go into the class and then how they feel after completing the semester. It made perfect sense." About whether or not his answers to the then-survey were consistent with how he felt at the beginning of the semester, Robert said, "I think so. I basically went into the class not liking math and I pretty much knew that. So, I think I answered pretty consistently." In each section below, I compare summaries of students' interview responses to their survey responses. I present pre-, post-, and then-survey results for math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness in Tables 17, 18, 19, 20, and 21, respectively.

Surveys vs. Interviews: Math Equanimity

Table 17

Interviewees' Math Equanimity Survey Scores

	Time		
Interviewees	Pre-Survey	Then-Survey	Post-Survey
Mia	2.5	3	3
Casie	3	1.75	3.25
Robert	1.75	3.25	2.75

Note. Equanimity scores range from 0 to 4, with 4 being associated with the highest amount of equanimity.

The interviewees' online DAPVU evidence marker scores for math equanimity are listed in Table 17. Mia, Casie, and Robert all described having test and performance anxiety that is not math-specific. Mia described how she had wrestled with general anxiety for a long time and how God helped her overcome much of her anxiety. She exhibited some math anxiety when I asked her about working through the think-aloud problems, but quickly agreed. Her survey scores were consistent with her claim that her math anxiety had not shifted much from the beginning to end of semester (increase of .5 points using the pre-survey, no change using the then-survey), and her scores reflected the levels of math anxiety she portrayed during the interview (slightly anxious).

Casie detailed several discouraging prior math experiences and was very math anxious at the start of Foundations. At the end of the semester, she had some lingering math-related anxiety, but the strength of her math anxiety had dramatically decreased. Her post-survey seems to accurately reflect her newfound math equanimity (score of 3.5 on a scale of 0 to 4), but her substantial positive change is only reflected using the then-survey (1.5 point increase). Her pre-survey and post-survey were only negligibly different (.25 point increase).

In addition to test and performance anxiety, Robert talked about his social anxiety (addressed below in the belongingness section) and anxiety from being spread too thin, with numerous obligations and deadlines. Robert had avoided math due to his math anxiety and he continued to have math anxiety at the end of the semester. He was most anxious when he had to work within certain time constraints or had to rely extensively on his memory (e.g., during tests), but his math anxiety did not appear to have an overall debilitating effect on his performance. I believe he slightly increased in math equanimity; the direction of this change agrees with the pre- and post- change scores (1 point increase), but disagrees with the then- and post- change scores (.5 point decrease).

Surveys vs. Interviews: Math Mindset

Table 18

Interviewees' Math Mindset Survey Scores

	Time		
Interviewees	Pre-Survey	Then-Survey	Post-Survey
Mia	5	5	5
Casie	5	5	5
Robert	2.67	1.67	3.67

Note. Mindset scores range from 0 to 5, with 5 being associated with the strongest agreement with a math growth mindset.

The interviewees' online DAPVU evidence marker scores for math mindset are listed in Table 18. All three students had a mix of performance and learning goals, but their learning goals outweighed their performance goals. When I asked each student if he or she believed there was such a thing as a math person and if he or she considered him or herself to be a math person, I wanted to uncover information about their math mindset as well as their math identity and interest. This phrasing may not have fully captured if they believe intelligence is malleable.

Mia attributed success to hard work and innate ability. She did not identify herself as a math person because she does not have a deep interest in math. She was mastery-oriented—she used effective strategies, monitored her progress, showed positive affect, and worked hard to accomplish her goals. Mia watched the mindset video (*How We Learn—Synapses and Pathways*) prior to taking the survey, and this could have impacted her pre-survey scores. When I asked Mia if she thought her then-survey responses on the mindset questions would match her pre-survey responses, Mia said, "I tend to be very positive now. So a lot of it was good [at the beginning] and it stayed that way [through to the end]." This positive attitude was reflected on her pre-, post-, and then-scores. She mentioned she was on the fence about whether or not everyone has the ability to do math, but said everyone is "able to learn the same things." This seeming contradiction stemmed from her belief that people have different innate abilities. While I believe Mia had a strong growth mindset, it may not have been as strong as her survey responses suggest.

Casie also watched the mindset video (*How We Learn—Synapses and Pathways*) prior to taking the survey, but she said it didn't stick at first. It wasn't until Frameworks that she really bought into the idea. I am uncertain as to whether Casie felt like it stuck prior to or after taking the pre-survey. Casie's adoption of a strong growth mindset was not reflected on her surveys because she maintained stable, high pre-, post-, and then-scores on the Math Theory of Intelligence Scale. Her post-scores agree with interview findings, but her identical pre- and then-scores are much higher than the interview suggests they should be.

Robert was unsure if the pre-survey was administered prior to or after covering the math mindset material. Robert exhibited some characteristics of mastery-oriented response patterns such as utilizing effective strategies, monitoring his progress, and persisting. However, he also showed some negative affect and was apprehensive about taking on new challenges. He believed he lacked innate math intelligence, though he did attribute his failures to lack of knowledge and not to a lack of innate math intelligence. When he said he does not consider himself a math person, I believe he was referring more to his lack of interest than his presumed lack of innate ability. Robert's many remarks about how persistence will lead to success indicate his mindset may have improved over the course of the semester.

Based on his survey responses, Robert cannot be labeled as either an entity or incremental theorist at the end of the semester, but his scores approach the threshold of an incremental theorist. I believe his post-survey is a true reflection of his end-of-semester mindset. Using either the pre-survey or then-survey, he is labeled an entity theorist at the beginning of the semester. The positive direction of survey change agrees with interview findings, but it is unclear if a 1-point change (using the pre-survey) or a 2-point change (using the then-survey) is more representative.

Surveys vs. Interviews: Math Self-Efficacy

Table 19

Interviewees' Math Self-Efficacy Survey Scores

	Time		
Interviewees	Pre-Survey	Then-Survey	Post-Survey
Mia	3.6	4	4
Casie	4	3.6	4
Robert	2	1.2	2.2

Note. Self-efficacy scores range from 0 to 4, with 4 being associated with the highest amount of self-efficacy.

The interviewees' online DAPVU evidence marker scores for math self-efficacy are listed in Table 19. At the end of the semester, Mia and Casie viewed themselves as capable to productively wrestle with Foundations math problems until they understood the content. As discussed in the literature review, attributions serve as a cue for efficacy appraisal. Mia and Casie predominately attributed their prior mathematical failures to lack of effort and their Foundations successes to sustained effort. They both had mastery experiences in Foundations. Casie experienced authentic, positive social persuasion from her instructor. Mia highly praised the instructor and made comments about liking the way he interacted with the students. I believe it is highly probable that Mia received encouraging feedback similar to what Casie experienced because they had the same instructor, but I have no means of determining whether or not that was the case.

Mia and Casie reported the highest level of self-efficacy on the post-survey, and this matches what I witnessed in the interviews. Mia's then-survey matched her post-survey, but her pre-survey was slightly lower (.4 difference). Her pre-survey is in the direction I expected. Her pre- and then-survey scores are slightly higher than one would expect from a person who said she did not have faith in her math ability prior to Foundations. However, this high pre-score could reflect her new personal identity and general confidence. Casie's pre-survey matched her post-survey, but her then-survey was slightly lower (.4 difference). Her then-survey is in the direction I expected. The differences in the pre- and then-survey scores seem somewhat small to be meaningful, but it cannot be ruled meaningless without additional information. Based on the interview, I assumed Casie's pre-survey and then-survey would have been much lower than they were.

Robert had low math self-efficacy at the beginning of the semester. Two semesters prior to Foundations, he had been shamed by a math instructor for his inability to provide quick recall of math facts. While he logically understood how fast someone could recite times tables is a poor measure of general math ability, this and other negative math experiences impacted his math self-esteem. Robert had witnessed others fail while he achieved mastery experiences through much hard work, and mastery experiences are most powerful when they occur under these circumstances. Robert's math self-efficacy was still low at the end of the semester, but it had noticeably increased. His low then-survey score (1.2) and mid-range post-survey score (2.2) coincide with my expectations based on his interview, but his pre-survey score (2) is higher than anticipated (especially in comparison to his post-survey score).

Surveys vs. Interviews: Math Belongingness

Table 20

Interviewees' Math Belongingness Survey Scores

	Time		
Interviewees	Pre-Survey	Then-Survey	Post-Survey
Mia	4	4	4
Casie	4	4	4
Robert	1	3	2

Note. Math belongingness scores range from 0 to 4, with 4 being associated with the strongest sense of math belonging.

The interviewees' online DAPVU evidence marker scores for math belongingness are listed in Table 20. Mia and Casie felt like they belonged and enjoyed interacting with their peers to tackle math problems. Mia did not specify whether or not her sense of math belongingness increased over the semester, but she noted that she was still trying to interact with people more. Casie's sense of math belonging improved considerably; she even added how the experiences would help her cooperatively solve problems in other situations. Mia and Casie reported the highest level of math belongingness on all three surveys (all 4s). Mia's survey results coincide with what I learned in the interview, but Casie's pre- and then-scores are higher than expected.

Robert's sense of math belonging was enhanced during the semester. His previous math experiences made him not want to share his ideas initially, but this changed as the semester progressed. His increased willingness to share his mathematical ideas and feel like those ideas were accepted was threatened by his belief that he was viewed differently because of his age. He remained a little distant from some students, especially those who did not take the class seriously, and his strongest sense of math belonging may have been tied to the instructor's acceptance of his ideas. According to Robert's post-survey, he would "sometimes" wonder if he did not belong in his math class, and I believe this perfectly reflects our discussion. According to his then-survey, his sense of math belonging decreased from "hardly ever" and this change is in line with his concerns about the age gap. According to his pre-survey, his math belongingness increased from "frequently" and this is in line with how he described his initial intimidation. Both are reasonable explanations; so, it is difficult to determine whether the pre- or then-survey was a better measure of Robert's beginning-of-semester math belongingness.

Surveys vs. Interviews: College Belongingness

Table 21

Interviewees' College Belongingness Survey Scores

	Time		
Interviewees	Pre-Survey	Then-Survey	Post-Survey
Mia	4	4	4
Casie	4	4	4
Robert	3	3	3

Note. College belongingness scores range from 0 to 4, with 4 being associated with the strongest sense of math belonging.

The interviewees' online DAPVU evidence marker scores for college belongingness are listed in Table 21. Mia, Casie, and Robert were excited to be back in college and seemed to think it was long overdue. They had very specific career aspirations and recognized that obtaining their degrees would help them achieve these goals. They were very interested in learning and identified with the college as an academic community where they could acquire the skills they need to be successful. In other words, college was exactly where they thought they belonged. Their responses on their pre-, post-, and then-surveys were stable, and the sense of college belongingness I ascertained from their interviews mirrored their surveys. Mia and Casie reported the highest level of college belongingness (4s). Robert had lower scores (3s), but this stemmed from him not feeling attached to one particular college; as long as it was a decent college with caring instructors, it was sufficient.

Online DAPVU and Think-Aloud for Interviewees

In general, students did not spend much time on the online DAPVU. Robert spent approximately 18.5 minutes, Mia spent a little over 57 minutes, and Casie spent a little over 10 minutes. Table 22 lists the amount of time the interviewees spent on each problem situation page. Because I was unable to see whether or not they were actually working on the problem situations for the entire amount of time, the minutes and seconds in the table only represent the amount of time each respondent spent on each particular page before they submitted their final responses (by clicking the "Next" button) on that page. Tables 23, 24, 25, and 26 list the scores interviewees received for the online DAPVU evidence markers for Accuracy, Representation, Language, and Depth, respectively.

Table 22

Interviewees' Time Spent by DAPVU Problem Situation Page

	Problem Situation Pages					
Interviewees	PS1	PS2	PS3	PS4 Intro	PS4	PS5
Mia	6:31	2:09	4:11	0:41	4:46	39:23
Casie	1:56	1:22	1:43	0:16	2:55	1:30
Robert	2:30	2:02	3:59	0:26	2:51	6:44

Note. These times represent the amount of time (in minutes and seconds) each respondent spent on the page that corresponds to each problem situation. The amount of time spent on the page does not imply the respondent was working on the problem the entire time. The Rugolia problem situation is presented on two pages. The first page (PS4 Intro) describes the Rugolian money system and the second page (PS4) presents a shortened version of the same information with questions for the respondent.

Table 23

Interviewees' Accuracy by Problem Situation

	Problem Situations			
Evidence Markers	PS1	PS3	PS4	PS5
Inaccurate/no reproduction of error pattern	R, M, C	M, C	R, M, C	M
Partially accurate	/	/	/	R
Fully accurate				

Note. A / indicates that the problem situation does not utilize the corresponding evidence marker.
Note. The students are denoted as R (Robert), M (Mia), and C (Casie).
Note. Robert did not provide a computation on PS3. Casie did not provide a computation on PS5.

Table 24

Interviewees' Representations by Problem Situation

	Problem Situations		
Evidence Markers	PS1	PS3	PS4
Digits and units	C		C
Letters only	/	/	M
Symbolic	R, M	M, C	
Other (nonsensical)	/	/	R

Note. A / indicates that the problem situation does not utilize the corresponding evidence marker.
Note. The students are denoted as R (Robert), M (Mia), and C (Casie).
Note. Robert did not provide a computation on PS3.

Table 25

Interviewees' Descriptive Language Use by Problem Situation

	Problem Situations	
Evidence Markers	PS2	PS5
No or inaccurate use of place value language	M, C	M, C
Accurate, non-specific place value language used to indicate a behavior		
Accurate, non-specific place value language to indicate a concept		
Accurate, specific place value language		R
Accurate, highly specific place value language	R	

Note. The students are denoted as R (Robert), M (Mia), and C (Casie).

Table 26

Interviewees' Depth of Analysis/Understanding by Problem Situation

	Problem Situations				
Evidence Markers	PS1	PS2	PS3	PS4	PS5
0 ~ No evidence of understanding			M	R	C
1 ~ Evidence of partial procedural understanding		M, C	C	M, C	M
2 ~ Evidence of partial conceptual understanding	C				
3 ~ Evidence of conceptual understanding	M				R
4 ~ Evidence of conceptual understanding		R			
5 ~ Evidence of conceptual understanding		/	/	/	/

Note. A / indicates that the problem situation does not utilize the corresponding evidence marker.
Scores from 3 to 5 range from conceptual understanding with a minor conceptual or computational error to conceptual understanding with no errors. A score of 4 on PS2, PS3, PS4, or PS5 indicates no errors.
The students are denoted as R (Robert), M (Mia), and C (Casie).
There was insufficient evidence to decipher Robert's Depth of Understanding on PS1.
Robert explicitly expressed confusion on PS3, but did not provide a computation.

Interview Think-Aloud

During the think-aloud portion of the interview, I tried to remain especially cognizant of how my presence was affecting each interviewee. I attempted to put the interviewees at ease by explaining the focus was on their thinking and not their correctness. For example, I told Casie: "I get a little uncomfortable when somebody watches me do work. So, again, I try to say this as much as possible. It's no pressure and I'm more interested in how you think about the problem than whether or not you get the right answer. In fact, off the top of my head, I couldn't tell you the right answer to this problem…A lot of, even incorrect, answers have a lot of correct thinking along the way, but that's not obvious when you get the result." Casie asked, "It doesn't have to sound like an actual sentence does it? Because I think it's just going to be thought." I reiterated thoughts were exactly what I was after, and I did not care whether or not she expressed her thoughts in full sentences.

I tried to keep things light while explaining reasons for the types of questions I intended to ask: “If there’s something where I’m not really sure what you’re saying, or I want to make sure I’m not misinterpreting, then I’ll ask you. So, if I ask, ‘What do you mean here?’ that doesn’t mean it’s wrong. It just means I want to make sure I understand because I can’t actually see inside your brain…believe it or not.” I reminded the interviewees about confidentiality while explaining how they would not be adversely affected by incorrect answers: “There’s no grade here. Nobody’s even going to know it was you when I write about it. I’ll be like, ‘Person 532 said…” I also explained they could discontinue solving problems anytime they wished to do so.

I occasionally checked in with the interviewees to make sure I was not hovering too much or making them nervous. All three respondents appeared relaxed and willing to share their ideas and frustrations. I offered them the choice to either think aloud while solving the problems or explain their thinking after working through the problems. They predominately chose to verbalize their thoughts while solving the problems but sometimes pondered silently. Mia was relaxed and said my presence did not make her feel uncomfortable. Casie said it felt weird to “[talk] about the problem[s] out loud.” She explained, “I’ve never done that before, so it feels weird. Not necessarily negative. It’s just kinda interesting.”

All three respondents thought the online DAPVU problems were unusual. Casie called the problems weird, but they intrigued her because they made her think. Mia, Casie, and Robert each mentioned something about it being more difficult to explain their thinking online than in person. Robert said, in reference to PS5, “If I was to look at it now, and I could tell you in person, that would be easier for me." He said the issue was not just explaining his thinking online, but trying to explain it in a way he thought I would understand without me being there to ask questions. I agreed and said, even if someone did something perfectly logical on the online assessment, it could be challenging to determine exactly what the person was thinking and what actions the person had taken. Robert and I spent a lot of time chatting[45] during the interview, so we ran out of time before I could ask him to look over all of the think-aloud problem situations. I asked him if he would be interested in looking at some of the problems and giving me his opinion. I explained that he didn't have to solve the problems, but I was interested in how he would approach the problems if he were to solve them; he agreed to provide feedback.

Below, I share the results of the think-aloud portion of the interview and compare their in-person responses to the online assessment. For simplicity, I will use the online DAPVU problem situation shorthand (e.g., PS1, PS2) to identify online DAPVU problem situations and TA to identify the Think-Aloud problem situations.

DAPVU Pat's Skiing Competition (PS1) and TA Space Shuttle

TA Space Shuttle is similar to the PS1, but the former involves subtraction and the latter involves addition. The quantitative representation for both problem situations is Familiar-Nonsystematic. The units on TA Space Shuttle are separated by colons and positioned from left to right as follows: hours, minutes, seconds, hundredths of a second. This can be represented symbolically as HH:MM:SS:CC. PS1 did not include hours and a

[45] As a funny side story: Robert and I had a lighthearted conversation with a security guard. When the security guard asked about the purpose of the interview and Robert said it was about math, she said, "Don't talk dirty to me." I jokingly asked if she wanted to work through some math problems and she said, "I will arrest you, woman."

full stop was used to separate the seconds and centiseconds (i.e., symbolically as MM:SS.CC).

Mia—Pat's Skiing Competition (PS1) and TA Space Shuttle

Mia explained she didn't like PS1 because she didn't know how to add minutes and seconds in the way it was presented. I asked what about PS1 made it difficult and she said, "I have never done that before." First, I was going to try to subtract it, but that didn't look right…Then, I added them, but that was weird, too." She said she was not sure when I asked if she thought she had obtained a reasonable or correct answer. I asked if she would like to try TA Space Shuttle, and she agreed. She was reading the problem silently, but verbalized "hundredths of a second." Before she began to work the problem, I asked why she had said this, and she responded, "I've just never seen it kinda like that…I think weird things are funny."

Mia recorded the two times as shown in Figure 11.

11:	26:	33:	07
8:	44:	35:	16

Figure 11. Mia's scratch-work on TA Space Shuttle (1)

Mia said, "Hundredths of a second add up…to seconds." Those are hours, minutes, seconds, hundredths of a second. I'm gonna add these together." More than once, she seemed to second-guess her decision to add: "But see it says how many times *elapsed between*?" After saying, "I don't know if I'm gonna do it right or not," she started adding from right to left. She recorded the result of her first addition (23) two rows below the 16 in the hundredths of a second column and again on the row just below 16. Then she recorded her second addition result (68) two rows below the 35 in the seconds column. She said, “60 seconds is a minute,” wrote 60 under the 68 she had just recorded, crossed out the 68, wrote 8 below the 60, crossed out the 60, and wrote 8 above the 68 that she had struck through. She said, “So we get 45 minutes” as she wrote 45 below the 44 in the minutes column. She recognized “60 seconds goes into a minute,” and regrouped 60 of the 68 seconds as 1 minute. When she added the minute to the minutes column, she wrote her result in an area inconsistent with her method because she had been recording what would be her final result just below the bottom addend. Her work is in Figure 12.

1 1:	2 6:	3 3:	0 7
8:	4 4:	3 5:	1 6
	4 5:	8:	2 3
		~~6 8~~	2 3
		~~6 0~~	
		8	

Figure 12. Mia’s scratch-work on TA Space Shuttle (2)

Mia expressed confusion about what she had in the minutes column by saying, “Well, no. That wouldn’t be right though, not if we’re adding them.” After she explained what she had done, she made a notable side comment: “I’m sure there’s a certain amount of hundredths of a second that goes into seconds, but I don’t know what that is.” She had recognized the mixed-grouping place value structure, but may not have been able to convert values greater than or equal to 100 from the hundredths of a second column to the seconds column because she did not associate 100 hundredths of a second with 1 second.

Mia believed her solution was incorrect because she had not done anything with the two left columns. She thought she would need someone to show her how to solve them problem. I asked if she had any ideas about what she could do with the 11 and the

26. I added, “It’s okay if you don’t. I’m just curious.” She said, “Maybe add the 26 with the 45 or subtract them. I know I would have to convert some things and then subtract it. Put hundredths of a second in the seconds and seconds in the minutes and then subtract after I get all that, but I’m not exactly sure how I would do that.” I ask what she meant by “convert” and she said, “changing hundredths of seconds to seconds and minutes.” I asked why she thought she would need to do that and she said, “I don’t know. I guess because I don’t think it would work if I just subtracted it. I could try it, though.” She then rewrote the problem. See Figure 13

In reference to her work in Figure 12 (above), she said, “I was adding these to try to convert hundredths of seconds to seconds and seconds to minutes.” I asked why she chose to subtract (Figure 13), and she said, “It’s a possibility it might work.”

10		12	
0 ~~0~~	12 5	~~2~~ 12	10
~~1~~ ~~1~~:	~~2~~ ~~6~~:	~~3~~ ~~3~~:	~~0~~ 7
8:	4 4:	3 5:	1 6
2:	8 1:	9 7:	9 1

Figure 13. Mia’s scratch-work on TA Space Shuttle (3)

Mia was unclear about whether she needed to use addition or subtraction to solve TA Space Shuttle. She thought subtraction would help her convert values to base-60. She did not know she could operate in base-ten in the hundredths of a second column, but she recognized the hundredths of a second column might involve a different base than the other columns. Her scores on PS 1 and TA Space Shuttle were identical. She received an Accuracy score of 0 and a Depth of Understanding score of 3.

Casie—Pat’s Skiing Competition (PS1) and TA Space Shuttle

Casie was initially nervous working on PS1. She said she “converted it wrong” because she sometimes forgets the number in the SS position must be less than 60, but the number in the CC position is not required to be less than 60. On PS1, she had added the seconds (53.67 + 50.54=104.21) and divided by 60 to “convert the result to minutes.” The CC amount did not need to be modified because it was less than 100. Her method gave her approximately 1.73 minutes, and she mentioned that she “accidentally rounded.” Then she added the other 4 minutes to get 5.73 minutes. After getting 5.73, she said she was thinking it meant 5 hours and 73 minutes. She converted 60 minutes to one hour and said the final answer was 6 hours and 13 minutes. She thought she could solve it correctly if she worked through the problem situation again after recognizing her errors. Referring to the CC amounts in PS1, she said, “There would be a little bit carried over, but that can go all the way up to 99, I think, instead of just the 60.” She added, “Time always messes me up.” Her recognition of her errors indicates she understood much more than she exhibited on the DAPVU.

When I asked Casie if she would like to do TA Space Shuttle, she said she “would like a second chance at that one.” She said, “We learned this in Foundations. It’s relative versus absolute…We would take the ending minus the beginning…It’s absolute change.” Then she arranged the two times to use a modified version of the traditional base-ten subtraction algorithm (Figure 14).

11:	26:	33:	07
8:	44:	35:	16

Figure 14. Casie’s scratch-work on TA Space Shuttle (1)

Casie said she could not subtract from the 0 and correctly subtracted one second and added 100 hundredths of a second to the CC group before subtracting the amounts (Figure 15).

		32	10
11:	2 6:	~~33~~:	~~0~~ 7
8:	4 4:	35:	1 6
			9 1

Figure 15. Casie's scratch-work on TA Space Shuttle (2)

Figure 16 shows how she regrouped within in the SS column and subtracted 5 from 12.

		2	
		~~3~~ ^{1}2	10
11:	26:	~~3 3~~:	~~0~~ 7
8:	44:	3 5:	1 6
		7:	9 1

Figure 16. Casie's scratch-work on TA Space Shuttle (3)

Up to this point, she had correctly calculated. When she tried to regroup 1 minute into the SS column, she did not treat it as 60 seconds—operating in base-10, she treated it as 100 seconds. Then she took 4 minutes from the remaining 5 minutes in the MM column. See Figure 17.

		12	
	5	~~3~~ ^{1}2	10
11:	2 ~~6~~:	~~3 3~~:	~~0~~ 7
8:	4 4:	3 5:	1 6
	1:	9 7:	9 1

Figure 17. Casie's scratch-work on TA Space Shuttle (4)

To complete the subtraction in the MM column, she attempted to regroup 1 hour. Again operating in base-10, she treated it as 100 minutes instead of 60. She also incorrectly subtracted 40 minutes from 120 minutes, recording 70 minutes in the MM column. See Figure 18.

		12	
0	5	~~3~~ ^{1}2	10
1 ~~1~~:	^{1}2 ~~6~~:	~~3 3~~:	~~0~~ 7
8:	4 4:	3 5:	1 6
2:	7 1:	9 7:	9 1

Figure 18. Casie's scratch-work on TA Space Shuttle (5)

After completing these calculations, she said, "Now I have to convert that to actual time. That's always my problem." Correctly operating in base-60, she subtracted 60 seconds from the SS column and added 1 minute to the MM column, arriving at 2:72:37:91. Then, she subtracted 60 minutes from the MM column and added 1 hour to the HH column, obtaining a final answer of 3:12:37:91. Casie was not fully confident about her final result. She said, "I don't know. I think, maybe. Times is bad, especially with that hundredths of a second." After she solved the problem, she explained how CC could not be more than 100 and the others could not be more than 60. She said, "Hours can go on forever. I guess, once you get to 24, you convert it to a day, but there are no days on this one."

Casie represented her results in fully symbolic notation and correctly verbalized the meaning of the notation. She recognized and utilized the mixed-grouping place value structure, adapting the traditional algorithm to regroup sometimes accurately and sometimes inaccurately across units. Casie was inaccurate on PS1, but exhibited partial accuracy on TA Space Shuttle. She received a her Depth of Understanding score of 3 on TA Space Shuttle but only a score of 2 on PS1.

<u>Robert—Pat's Skiing Competition (PS1) and TA Space Shuttle</u>

Robert said the Space Shuttle problem situation looked the easiest out of the TA problem situations because it "looks like simple math, [like] subtraction." This surprised me because, when I had previously asked Robert about PS1, he said he did not think he had correctly solved it because of the "conversion factors." After I asked what he meant by "conversion factors," he said, "If I recall…I'm assuming it was you going down the hill at a certain speed, how long would it take you and there's different conversion factors on how to break that down into the units you're trying to get. It could be that they were going down in kilometers and they want you to switch that to miles per hour and I probably just glossed over that and took a guess to be honest." Later in the interview, I redirected Robert back to TA Space Shuttle and asked for his reaction. After he worked through TA Space Shuttle, I asked him to remind me whether or not he liked the similar PS1. He said he liked TA Space Shuttle "because [he] could kinda talk through it with someone" and this was easier than "doing it by himself."

Explaining TA Space Shuttle, he said, "You would just take the two numbers and subtract one from the other. Your total would be the difference between the two." I asked if he would perform the calculations in his head and he said he would write them down because of the length. I said he was welcome to write the problem. He said, “I guess you would have to because it’s going into minutes. Minutes would translate different into…So, hours, minutes, seconds, hundredths of a second.” I offered him the Echo Smartpen to solve the problem. He lined up the numbers vertically in standard algorithm format and labeled them as shown in Figure 19.

Hour	Min	Sec	Hundredths
11:	26:	33:	07
8:	44:	35:	16

Figure 19. Robert’s scratch-work on TA Space Shuttle (1)

He said 8:44:35:16 “would have to be the bottom number because [11:26:33:07] is bigger.” When I asked how he chose which number should be on top, he explained it actually did not make a difference, but performing the calculations with 11:26:33:07 on the bottom would result in a negative number. He then explained the difficulty of the problem was the “hundredths of a second.” He continued, “How many hundredths of a second are in each second? I don’t know that. I don’t know that.” Referring to the seconds column, he asked, “How can I borrow from minutes?” Correcting his question, he said, “No these are seconds. So, for every minute is 60 seconds. So, 60 plus. That would be 93. That would make that a 25….That would be a minute.” His starting scratch-work is replicated in Figure 20.

Hour	Min	Sec	Hundredths
	25	93	
11:	~~26~~:	~~33~~:	07
8:	44:	35:	16
		60	

Figure 20. Robert’s scratch-work on TA Space Shuttle (2)

He realized the 60 seconds he recorded indicated he had done something incorrectly. Noticing he also skipped the CC column, he asked if I knew how many hundredths of a second are in a second. After me reminding him I was not going to assist because I wanted to know how he was thinking about the problem, he said, “I’m gonna guess a 100 because it’s hundredths. I’m gonna say 100. I’m gonna start all over. I’m gonna say 107. He regrouped 1 second into the CC column, used a calculator to subtract 16 from 107, and regrouped 1 minute into the SS column, saying, “I borrowed off of [33], so that would be 32. I’m gonna borrow off of [26], so that would be 25.” See Figure 21.

Hour	Min	Sec	Hundredths
		92	
	25	~~32~~	107
11:	~~26~~:	~~33~~:	07
8:	44:	35:	16
			91

Figure 21. Robert's scratch-work on TA Space Shuttle (3)

I asked how he arrived at the 92 and he said, "I got the 92 because this I turned into a 32 so I took one second off. And then there are 60 seconds in a minute. So, 60 + 32 would be 92." He subtracted 35 from 92 and placed 57 under the subtrahend in the SS column, and said, "That's much better." I asked, "Why is that better?" and he responded, "The way I had it before it was 60 seconds and that would have made a minute. So, I would have converted it back over. So, I could tell something was wrong in the equation if that was the case." In other words, he realized if the result of the appropriate subtraction were actually 60 seconds, he would not have needed to regroup initially. He then explained he needed to "borrow" from the HH column to add 60 minutes to the MM column, writing 85 at the top of the MM column. He completed the calculation by subtracting 8 from 10. See Figure 22.

Hour	Min	Sec	Hundredths
	85	92	
10	~~25~~	~~32~~	107
~~11~~:	~~26~~:	~~33~~:	07
8:	44:	35:	16
2:	41:	57:	91

Figure 22. Robert's scratch-work on TA Space Shuttle (4)

He concluded, "There's 2 hours and 41 minutes 57 seconds and 91 hundredths of a second." Robert accurately solved TA Space Shuttle, representing his results in fully symbolic notation and correctly verbalizing the meaning of the notation. He initially did not recognize that there are 100 hundredths of a second in a second but quickly deduced this. He recognized and utilized the mixed-grouping place value structure, adapting the traditional algorithm to regroup accurately across and within units. Robert received a top Depth of Understanding score (5) on TA Space Shuttle. On PS1, he received a score of 0 for Accuracy for providing an incorrect response, and I was unable to determine his solution method to assess his depth of understanding.

DAPVU Bobby's Squares (PS2)

The quantitative representation for PS2 is Familiar-Systematic. I did not ask the interviewees to think aloud about any problem situations similar to PS2. PS2 asks respondents to analyze Bobby's thinking. Bobby looked at a drawing of 26 squares and expressed the amount numerically (as 26). When Bobby's teacher asked him to circle the squares that were represented the 6, Bobby circled 6 squares; but, when she asked him to box the squares represented by the 2, he only boxed 2 squares. Casie brought up an issue with PS2: The numbers 2 and 6 in the 26 were typed further apart than intended, and Casie attributed Bobby's misunderstanding to this formatting error. Another source of possible confusion, not mentioned by Casie, is the squares were presented in such a way that a "box" could not be drawn around 20 squares. See Figure 23.

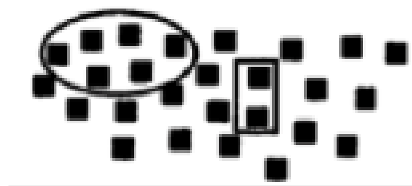


Figure 23. Portion of Problem Situation 3: Bobby's Squares

<u>Mia and Robert—Bobby's Squares (PS2)</u>

I did not have sufficient time to discuss Bobby's Squares with Mia or Robert. For Mia's online responses, she received a Descriptive Language score of 0 because she did not use place value language and a score of 1 on Depth of Analysis because she described some behaviors, but no analysis of understanding. Online, Robert articulated Bobby's lack of association between the digit 2 and two groups of 10. He received a Descriptive Language score of 4 and a Depth of Analysis score of 4.

<u>Casie—Bobby's Squares</u>

As mentioned above, Casie thought Bobby's confusion stemmed from the spacing of the 2 and 6 in 26; she thought Bobby did not recognize the 26 as "one whole number." In the problem situation, Bobby was the one who had written the 26, but Casie mistakenly assumed that the teacher had written 26 strangely. This was likely due to the fact that Casie had spent very little time on PS2. Explaining Bobby's thinking further, she said, "Either way, he's got 6 circled and he's got 2 circled. So he's got the 6 and the 2. He just hasn't put it together as a whole number yet. So he's just got two separate numbers that doesn't add up to the actual number. But I see what he's doing. I probably did stuff like that as a kid too. It's so hard to explain stuff like that." In her interview, Casie evidenced a Descriptive Language score of 1 on PS2 because she used non-specific place value language in reference to Bobby's behaviors. This score is 1 one point higher than on her online responses. She received a Depth of Analysis score of 1 for both her online and interview responses regarding PS2.

DAPVU Chocolate Factory (PS3) and TA Chocolate Factory

The quantitative representation for TA Chocolate Factory and PS3 is Unfamiliar-Systematic. Factory works in the scenarios use base-four notation to represent how many chocolates are in differently sized receptacles (four singles complete one package, four packages complete one box, four boxes complete one carton, and four cartons complete one case). TA Chocolate Factory requires the subject to represent 30 packages of chocolate in factory notation (i.e., 1320). PS3 requires the subject to subtract 3012 from 10000 using factory notation (i.e., 10000 (base-four) – 3012 (base-four) = 322 (base-four)).

<u>Mia—Chocolate Factory (PS3) and TA Chocolate Factory</u>

Mia said she remembered exactly what she had done on PS3; She meant to subtract 3012 from 10,000, but instead subtracted 1,000 from 3012. Even if she had performed the calculation she intended, her answer would have been incorrect because she was operating in base-ten. During our discussion, she said she, initially, must not have considered the problem closely because she had not used the listed equivalencies (e.g., 4 packages = 1 box) when solving the problem online. However, she claimed she understood PS3 after reviewing it further. I asked how she thought her previous interpretation might have affected her result. She said she was uncertain but she could try the problem situation again with new insight.

Mia began PS3 during the interview by correctly denoting 3012 as 3 cartons, 0 boxes, 1 package, and 2 singles. She alternated back and forth between believing she did and did not understand PS3 during the interview. She said it was confusing because she did not "know how to work it because [she had] not been instructed on anything like [PS3]." She eventually said, "They would need 2 more singles to make 1 package." After rereading the problem situation a few times, she said, "I really don't know." I asked, "What did you do here?" in reference to the line of her scratch-work that said, "2 single choc + 2 singles = 1 package." She explained, "I know that 2 more singles would make a package and then I would have to make more packages to get more boxes and more boxes to get more cartons, all the way up to the thing, but I don't know how I would do that exactly." I asked if she thought it would be easier if I were not sitting there, watching, and she said, "After a while. It might take some time to get it."

She agreed to try TA Chocolate Factory and started by implying the TA problem situation may be easier because "it's a whole number." She divided 30 by 4 and, after looking at what she had done and rereading the problem, she said, "I don't know. I just don't know how to do it, I guess…I would have to make a table for this one, but I would have to remember how to do that." I was unsure as to what kind of table she thought would help. When I asked, she explained she has a bad memory and just did not remember how to do the table. She kept going back to her division (30 ÷ 4), but she was unsure about its meaning and eventually said, "I really don't know how to do it. Is that okay?" I assured her it was fine and we could stop whenever she wanted to stop; she said she wanted to discontinue working on TA Chocolate Factory. I inquired if working on the problem situation was stressful. She said, “No, it’s not stressful. I just don't want to do it.” However, she was willing to look at other problem situations. Mia received a 0 for Accuracy for PS3 online. For her PS3 and TA Chocolate Factory interview responses, she received a 99 (no computation) for Accuracy. Online, she had not recognized the base-four place value structure of PS3 and received a score of 0 for Depth of Understanding. During her interview, she recognized the structure, but was unable to apply appropriate strategies, evidencing a Depth score of 1 for PS3 and TA Chocolate Factory.

Casie—Chocolate Factory (PS3) and TA Chocolate Factory

Casie and I did not discuss PS3 in depth and we were unable to discuss TA Chocolate Factory. Casie spent more time on PS3 online than she believed to be appropriate (only 1 minute, 43 seconds). Casie said relating PS3 to something familiar assisted her in solving the problem. She compared PS3 to “dollars and cents, but with four singles worth a dollar.” She continued, “We’re just breaking it down a little bit and then each one adds up to the next one like those little [Russian] dolls. One case is the 4 cartons and there are 4 cartons in the…and I was, ‘Okay, okay, I get that.’ It’s all inside each other. It’s all the same amount in different areas.”

I believe Casie arrived at her online answer through a calculation error and a misconception. It appears as though she tried to subtract 3012 (the amount in the partially full case) from 4444 (instead of subtracting it from 10000). Because she made a calculation error, her result was 1032. Then, she converted 1032 (base 4) to 78 (base 10) and wrote, “They need 78 more chocolates.” She received an Accuracy score of 0, a Representation score of 1, and a Depth of Understanding score of 1 on her online solutions to PS3.

Robert—Chocolate Factory (PS3) and TA Chocolate Factory

Robert said, if he were to attempt all the TA problem situations, he would save TA Chocolate Factory for last because it would be the most difficult and take the longest amount of time. He drew comparisons between TA Chocolate Factory and TA Rugolia, but said TA Rugolia would be easier because it has "more visual aids" than TA Chocolate Factory. Robert explicated why he thought TA Chocolate Factory would be difficult: "There are so many words in that and numbers…It's four singles, four packages, four cartons. There's no pattern. There's nothing. It's just 4, 4, 4 and you've gotta convert it…It's very easy to get jumbled up. You'd have to really sit down and write it out, at least for me." I asked if he had written out PS3, and he responded, "Probably not. I probably guessed at it because I didn't care. Sorry." I reassured him that I was happy he was being honest. He said, "Yeah, I'm just looking at [TA Chocolate Factory] and it makes my head hurt…For me, it's difficult because, at least on [TA Rugolia], I can see—I have visual aids which makes it easier for me to do. If I'm just straight thinking about numbers, it's a little more difficult."

I asked Robert how he would approach TA Chocolate Factory if he were to try attempt it, and he said, "I would probably have to break down each one. So, 4 singles equals 1 package. So, 4 packages would be 4 times 4. That would be 16 singles per 4 packages. Then for 1 box it would be the same. You just kinda crisscross on the multiplication." He added, "I still don't like it. I still don't like it." It was obvious, in person, that Robert did not understand the chocolate factory notation; he said the notation "means number of cases times number of cartons times number of boxes times number of packages times number of singles." I am not sure what he meant by "there's no pattern" because he clearly recognized the base-four place value structure and described regrouping strategies.

Robert did not try to solve TA Chocolate Factory nor did he provide a numeric response to PS3. He initially viewed PS3 on his phone and was having difficulty because not all the information fit on the screen. In his online response, he said he was unsure about how to solve the problem and he "would really have to sit down and break down the amounts to understand the problem." Because the Accuracy, Representation, and Depth of Understanding evidence markers rely on problem attempts, I was unable to assign meaningful evidence scores for TA Chocolate Factory or PS3. However, Robert evidenced greater depth of understanding about the Chocolate Factory problem situations during the interview than in his responses to PS3 online.

DAPVU Rugolia (PS4) and TA Rugolia

TA Rugolia and PS4 have the same setup, but the former involves multiplication and the latter involves addition. The quantitative representation for both problem situations is Unfamiliar-Nonsystematic. In the Kingdom of Rugolia, Y is the basic coin, G is worth 4 Ys, R is worth 2 Gs, and B is worth 3 Rs. In PS4, the respondent is asked to add two prices (GYYRBRYRBRY and GYBBGYYGBGY). In TA Rugolia, the respondent is asked to represent the cost of 5 rugs, where each rug costs RRGYY, using the fewest number of coins possible.

<u>Mia—Rugolia (PS4) and TA Rugolia</u>

Mia chose TA Rugolia as her first TA problem situation. She explained what she was doing as she worked through the problem, and she arrived at her answer very quickly. She created 5 rows of RRGYY and crossed out letters as she converted the coins to larger denominations. See Figure 24.

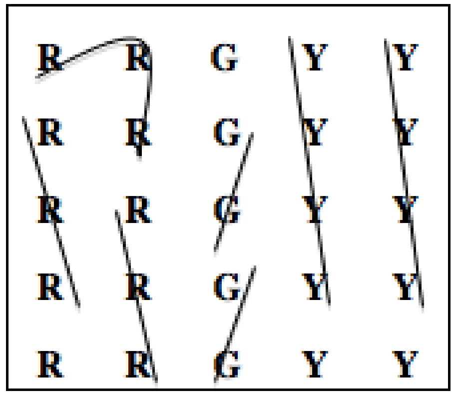


Figure 24. Mia's scratch-work array on TA Rugolia

On the first line under her scratch-work array, Mia wrote GG to represent the 2 sets of Ys, RRR to represent the 2 sets of Gs and the R she had not crossed out, and YY to represent the Ys she had not crossed out. The expression was: GGRRRYY. She had not recorded the 3 sets of Rs or the 1 G she had not crossed out. On the second line below her array, she wrote BBB to represent the 3 sets of Rs. She crossed out the 3 Rs in line 1 and recorded an additional B. Her expression on line 1 was GG~~RRR~~YY and her expression on line 2 was BBBB. She then crossed out 1 G in line 1 (i.e., G~~GRRR~~YY) and recorded 1 G in line 2 (i.e., BBBBG). She used the G she had not crossed out and the remaining G in line 1 to add an R to line 2 (i.e., BBBBGR). She completed the problem by adding the leftover Ys from line 1 to line 2. Her final answer on TA Rugolia was BBBBRGYY, the correct price of the rugs.

Mia said it was simpler for her to work the problem by exchanging letters than by using a table. Mia recognized the mixed-grouping place value structure on TA Rugolia. She applied regrouping strategies to represent the quantity, using the fewest number of coins possible, without computational errors. She said she applied the same regrouping strategies to the mixed-grouping place value structure on PS4. However, her response online to PS 4 was incorrect. She received a Depth of Understanding score of 1 on PS4 and exhibited a Depth of Understanding score of 4 on TA Rugolia.

Casie—Rugolia (PS4) and TA Rugolia

Casie and I were unable to talk about TA Rugolia, but she briefly provided her thoughts about PS4. She believed she might have done well on PS4. She said, "it took me a little bit of time, but I was like, 'This one kinda makes sense.'" She said it reminded her of counting Legos as a kid because of all of the different colors and sizes and how she would stack them up by color. I asked how it related to the differently sized Legos. She said, "I had to write all that down while I was going, but the Legos, I think were little size or something. We kinda used it like trading, but [that was] 20 years ago."

I am unsure how Casie chose the initial amounts for the 2 rugs. She said her starting amount was 8Y, 7G, and 7R. Using her initial amounts, she correctly regrouped in her first step (8Y=GG, 7G=RRRG, 7R=BBR). This result should be converted to 3B, 2R, and 1G; her final answer was 2B, 1R, and 1G. She received an online score of 0 for Accuracy and a score of 1 for Depth of Understanding because she recognized the mixed-grouping place value structure but inaccurately applied regrouping strategies.

Robert—Rugolia (PS4) and TA Rugolia

Robert had stopped working on the online DAPVU when he was trying to decipher the Rugolia problem because he was distracted and in a poor mood, but he returned to it later. He did not use paper while solving PS4 online; he "counted them up in [his] head as [he] went along." Robert also solved TA Rugolia by keeping "a calculator" in his head. Robert misinterpreted the problem situation setup for PS4 and TA Rugolia. At first, I had a very difficult time figuring out how he arrived at his answers. The prices of the rugs in TA Rugolia are given, but Robert said he would "find out what each rug costs." His response to TA Rugolia was 10 and his response to PS4 was 22. His online description of his solution method was, I counted out all the coins, but from looking at the color green, I was able to tell it would use less coins." Robert interpreted the problem setup as: Y=1, G=4, R=2, and B=3. In TA Rugolia, the respondent is supposed to multiply the cost of one rug (RRGYY) by 5. Using Robert's method, the rug is worth 2R + 1G + 2Y = 2(2) + 1(4) + 2(1) = 4 + 4 + 2 = 10. Robert received an Accuracy score of 0 and a Depth of Understanding score of 0 on PS4 and TA Rugolia.

DAPVU Maria's Subtraction (PS5) and TA Mark's Addition

TA Mark's Addition is similar to PS5, but the former involves addition and the latter involves subtraction. The quantitative representation for both problem situations is Familiar-Systematic. TA Mark's Addition and PS5 ask the respondent to replicate a hypothetical student's error patterns on 3 problems. The results for correctly reproducing the error on TA Mark's Addition are 193, 867, and 1116. The results for correctly reproducing the error on PS5 are 212, 356, and 266.

<u>Mia—Maria's Subtraction (PS5) and TA Mark's Addition</u>

Mia focused on Mark's behaviors, not on his conceptual understanding. She said, "He's working from left to right instead of right to left. He's just working backwards. After Mia described Mark's lack of procedural understanding, I turned Mia's attention to where the problem situation asks for a description of Mark's conceptual understanding. Again, she focused on behaviors, saying, "He doesn't understand you are supposed to go from right to left…He's putting the wrong one up there. You're supposed to put the first number up here and he puts the second number up here, but he's completely backwards." Hoping to elicit place value language, I asked, "What do you mean by the first number and the second number?" She pointed at the number in the ones column and said it should be placed below the addends and pointed at the number in the tens column and said it should be placed above the addends; she did not make any references to “tens” or “regrouping”. After additional probing, she said, “He’s carrying the wrong number, but he’s completely backwards.”

Mia and I did not have the opportunity to discuss PS5, other than her saying, “I just worked them out to see what [Maria] was doing.” Mia did not correctly reproduce Maria’s error pattern on PS5, but she did correctly reproduce Mark’s error pattern on TA Mark’s Addition; hence, she received an Accuracy score of 0 on the former and 2 on the latter. She received a score of 0 for her Use of Descriptive Language on PS5, but a score of 1 on TA Mark’s Addition. The 1-point increase was due to her use of the word “carrying” after my extensive probing. She evidenced a score of 1 for Depth of Analysis on PS5 and TA Mark’s Addition.

<u>Casie—Maria’s Subtraction (PS5) and TA Mark’s Addition</u>

Talking about PS5, Casie said, “I stared at that and I just kinda kept staring at it and eventually I had to answer with, ‘I have no idea what she is doing.’ I had no idea.” She looked at PS5 again and said, “No matter how much I look at it I just I don’t know what she’s doing…She just completely didn’t take a number off that one…It’s all weird. I just don’t understand it at all. I’m just so lost.”

Casie accurately reproduced Mark’s error pattern on TA Mark’s Addition. She recognized why Mark would have solved one of the problems correctly and said, “That one, since there’s no carrying, [he was correct].” She consistently focused on Mark’s procedural errors and explained the impact of those errors on his solutions: “Carrying the number over to the right side kinda messes up the whole problem…You just end up with these weird answers.” She offered a hypothesis about why Mark may calculate in this manner, saying, “I could see why he would do that if no one ever taught him you have to go from right to left because that’s how you read English.” I asked why “it doesn’t work from left to right,” and she said, “It’s because you have to carry the number from the right to the left…I’m not sure why. I’m sitting here trying to figure out why it is we have to do that. The only answer I have for that is, ‘if you do it that way, it’s wrong.’” She talked about how he “carries the second half of the number instead of the first half.” I asked what she meant by “the second half” and “the first half” hoping to elicit a response involving specific place value language and a deeper analysis of Mark’s thinking. Instead of referring to ones or tens, she again described the process. Casie thought Mark’s methods were interesting and said she was “having fun trying to figure out” how the standard algorithm works.

Out of all the TA problems, Casie said she was most confident on TA Mark’s Addition. Out of all online DAPVU problems, she had been least confident on PS5; she said it was much easier to find Mark’s addition error pattern than Maria’s subtraction error pattern. Casie did not provide a numerical response online to PS5. She received an Accuracy score of 99 (missing response) on PS5 and a score of 2 on TA Mark’s Addition. During the interview, she used accurate but non-specific place value language to describe Mark’s behaviors, but she provided no analysis of Mark’s conceptual understanding. She received a score of 0 for both the Descriptive Language and the Depth of Analysis evidence markers on PS5. She received a score of 1 on the same evidence markers for TA Mark’s Addition.

Robert—Maria’s Subtraction (PS5) and TA Mark’s Addition

During the interview, Robert described PS5 as the problem with “the girl who was doing borrowing wrong.” He said, “I could see what she was doing wrong, but it was hard for me to put it into words. Recalling PS5, Robert said, "There were some of them where she would subtract and she was taking out of the tens column but she wasn't actually crossing out and taking one of the units off. She was just borrowing straight from the two and carrying over without giving the correct measurement, units, in the tens column." I probed further to determine if he thought Maria's errors were due to inappropriate technique or lack of understanding and he said, "She did not understand." At this point, a security guard interrupted us; we did not go back to PS5, and we did not discuss TA Mark's Addition.

Robert was partially accurate in his reproduction of Maria's error pattern PS5 online. Robert used accurate and specific place value language in person and online, though some of it was informal (e.g., talked about the "tens" column instead of "groups of tens"). He received a score of 3 for Descriptive Language on PS5 for both his in-person responses and his online responses. He developed an accurate analysis of the place value concepts Maria did not understand, and he received a score of 3 for Depth of Analysis.

DOCUMENT ROADMAP

In Chapter 4, I further described my key variables and measures. I discussed how my data met the assumptions of multilevel models, and I provided quantitative results for Research Questions 1A, 1B, and 2. In an attempt to address Research Question 3, I explained my findings from responses to the online DAPVU through the lens of evidence markers. I used my one-on-one interviews to establish a clearer picture of the findings from my quantitative analyses of Research Questions 1A, 1B, and 2, and qualitative analysis of Research Question 3. Specifically, I detailed reflections and beliefs interviewees expressed about Foundations and how it impacted their dm-noncognitive factors, and I compared interviewees' comments to their pre-, post-, and then-survey scores. I described the interviewees' thought processes and actions for select DAPVU problem situations and similar think-aloud problem situations, and I compared their in-person responses to their online DAPVU responses. In Chapter 5, I review the aims of my research and paint a more cohesive picture of what I have learned through this extensive process. I outline the implications of this work for my future research, as well as potential implications for practitioners and the mathematics education research community at large.

Chapter 5: Discussion

I began this work in an effort to answer several questions related to my deep interest in the success of community college students, and I continued over a path that would have been hard to predict. This path led me ultimately to four questions that became the base of this study. The journey has taught me a great deal about the realities of research and produced a long list of things that, if I were to start again, I would do differently. Nonetheless, it has been quite a ride and I hope to put that knowledge to use in my subsequent work.

This chapter is broken into six parts. First, I review the study objectives and study elements. Next, I discuss the results in the context of prior research. Third, I describe the study complications and limitations. Then, I talk about how my dissertation contributed to the body of knowledge in the field. In the fifth section, I share my future research plans. I end the chapter with an overall summary and indicate the relevance of my dissertation research.

Review of Objectives and Study Elements

In very brief summary: My hopes for my dissertation study were to learn if developmental mathematics students enrolled in Foundations of Mathematical Reasoning would exhibit positive changes in select developmental mathematics noncognitive factors—math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness—and if these changes would be associated with their semester outcomes. I wanted to discover whether or not they would evidence their ability to transfer their knowledge to novel place value situations. I aspired to add to existing research about time- and cost-effective data collection measures that could potentially decrease the threat of response shift bias on self-report surveys.

I measured change using the self-report pre-post-then-surveys of Foundations students' math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness. To address my research question about changes in dm-noncognitive factors from the beginning to the end of the semester (RQ1A), I compared Foundations students' pre-surveys to post-surveys. To address my research question related to the utility of retrospective pretests (RQ1B), I compared pre-surveys to then-surveys. I collected students' course grades, final exam grades, and percent attendance, and I compared these outcomes to pre-post-survey change scores to address my research question concerning the relationship between changes in dm-noncognitive factors and semester outcomes (RQ2). I used the online Developmental Assessment of Place Value Understanding to address my research question about concept transfer (RQ3). I analyzed Research Questions 1A, 1B, and 2 using multilevel models, and I analyzed Research Question 3 by creating and assigning evidence markers to students' responses. I also interviewed three students about their dm-noncognitive factors and experiences in their Foundations course and asked them to think aloud on place value problem situations to help explain my findings. In the next few sections, I provide a high level summary of the results from my analyses.

SUMMARY DISCUSSION OF RESULTS

Question 1A: Pre to Post Changes in DM-Noncognitive Factors

I hypothesized students enrolled in Foundations of Mathematical reasoning would exhibit positive changes in their dm-noncognitive factors from beginning to end of semester. Based on Foundations students' self-report pre- and post-surveys, students were more math equanimous at the end of the semester than at the beginning of the semester.

This change cannot be attributed solely to the Foundations course because I had no comparison group. None of the control variables in the models used to address equanimity were significant, indicating that this change may be, at least partially, influenced by Foundations. My quantitative analyses suggested students did not experience changes in the other four dm-noncognitive factors. From the beginning to end of semester: Students mostly disagreed with the entity statements about math intelligence; they hovered between feeling halfway to mostly math self-efficacious; and they rarely to sometimes felt a low sense of math or college belonging.

My qualitative results more closely align with my gut feeling about the impact of Foundations and quantitative and qualitative findings of others (Bryk et al., 2013; Yeager, Bryk, et al., 2013; Yeager, Paunesku, et al., 2013). My interviewees all praised the course and course instructor. They had all experienced math struggles in the past, but they appreciated the utility of math after Foundations and mostly described positive changes in their dm-noncognitive factors. They discussed their willingness to persist, even in the face of failure. Casie exhibited the most positive changes, followed by Mia, and then Robert. Some of their survey results did not match their interview responses. For instance, Casie detailed how her math sense of belonging changed drastically during the semester, but she reported the exact same score on her pre-survey as her post-survey. At both time points, she reported the highest level of math belongingness. Lack of significant change on some of these dm-noncognitive factors could be due to a ceiling effect.

I cannot make any generalizations based on my interviews with three students. In addition to this being a small sample size, two of the students took Foundations with the same instructor (albeit in different sections), and their changed dm-noncognitive factors could have been influenced greatly by the instructor. Mia, Casie, and Robert attributed some of their changes to the instructors and the instructors' teaching methods. They provided examples of ways their instructor engaged the students, encouraged group work, and made math relevant to their personal lives. Casie said many wonderful things about her instructor, including that he was the best teacher she had ever met. It is possible that Mia and Casie's instructor and Robert's instructor would teach in the same manner (as described by the interviewees) in a different math course.

However, evidence suggests Foundations contributed to their positive change. First, many of the examples the interviewees gave about the instructors' methods are built into the Foundations curriculum. For example, the curriculum is set up to promote math belongingness and many of the lessons include legitimate real world situations. Second, Robert spoke to the differences between his Foundations course and a math course he had taken previously with the same instructor. He discussed how the instructor used two different modes of interaction in the separate courses.

Question 1B: Pre to Then Changes in DM-Noncognitive Factors

I hypothesized Foundations students would exhibit response shift on a self-report then-survey of math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness. I believed differences between pre- and post-survey scores on these dm-noncognitive factors would be smaller than differences between then- and post-survey scores.

Then-survey reports were typically higher than pre-survey reports, though response shift was only significant for math equanimity. I used a very conservative alpha (.01) for this research question. With alpha set at .05, math mindset, math self-efficacy, and math belongingness were still not significant. College belongingness had a p-value of

.022, and students exhibited response shift with then-scores significantly lower than pre-scores. Because this is an exploratory study, I will discuss college belongingness as if it were significant in the remainder of this section.

According to Hill and Betz (2005), proponents of true pretests and proponents of thentests agree that pre-post change scores are typically smaller than then-post change scores. My study findings for math equanimity run contrary to this, but my findings (with a less restrictive alpha) for college belongingness agree with this. One of the biggest concerns of response shift theorists is pre-post change scores underestimate program effects, while advocates of true pretests claim then-post change scores overestimate program effects. In no case did using a retrospective pretest in this study result in bigger program effects for the significant factors. It was quite the opposite—there were no significant differences between then- and post-scores.

Even if there had been response shift on all of the dm-noncognitive factors, I would not be able to say definitively that the shifts were due to response shift bias because I did not have access to a control group, a measure of nonsubjective change, a memory thentest, or a sufficient number of interviews. Below, I compare how differences between pre- and then-surveys might be interpreted through the lenses of the three competing theories I discussed in the literature review: response shift bias theory, personal recall theory, and impression management theory.

Response Shift Bias as an Explanation: Students were actually more math equanimous at the beginning of the semester than they reported on their pre-surveys. At the end of the semester, they responded to the then-survey with a recalibrated internal metric. They realized they had underestimated their initial equanimity, and they reported their true beginning-of-semester equanimity on the then-survey. Students maintained stable levels of math equanimity throughout the semester. Students overestimated their sense of college belonging at the beginning of the semester. Looking back, they realized they felt more like an outsider than they initially reported and adjusted their scores on the then-survey to represent their true beginning-of-semester college belongingness. Even with the more accurate measure (then-survey) of their initial sense of college belonging, their sense of college belonging did not improve over the course of the semester.

Personal Recall Theory as an Explanation: Students who could not recall how they felt at the beginning of the semester benchmarked their current state on the post-survey and then guessed at their beginning state. Foundations students adopted an implicit theory of stability about their math equanimity, and they were incorrect. They underwent definite changes during the semester, and this was evidenced using pre- and post-survey change scores. As Ross (1989) cautioned may sometimes be the case when an intervention brings about true change, they exaggerated the similarities between their beginning-of-semester and end-of-semester levels of math equanimity. The pre-survey was a more accurate measure of their initial math equanimity. Foundations students adopted an implicit theory of stability about their college belongingness, and they were correct. It is not possible to determine if the pre-survey or the then-survey was a stronger measure of beginning-of-semester college belongingness.

Impression Management Theory as an Explanation: Foundations students exhibited a consistency effect with their math equanimity and college belongingness post- and then-survey scores. Students generally could not recall their math equanimity or college belongingness scores. They inferred that it is a normal, and socially acceptable to be moderately anxious in a math course and they reported their math equanimity on their then-survey consistently as such. The pre-survey was a more accurate measure of their math equanimity. They recognized higher college belongingness scores were more socially desirable and consciously or subconsciously biased their then-surveys to appear as though they had always held a high sense of college belonging. It is not possible to determine if the pre-survey or the then-survey was a more accurate measure of beginning-of-semester college belongingness.

As noted above, I cannot say definitively which theory accurately explains the Foundations students' response shift on math equanimity and college belongingness. Pre-post pairwise comparisons of math equanimity scores showed significant positive change, but then-post pairwise comparisons were not significant. This suggests either the true pretest or the retrospective pretest is more accurate. Based on my interviewees' math equanimity responses, the pre-survey may be accurate for Mia and Robert, but the then-survey appears to be more accurate for Casie. Mia's responses seem to support a theory of stability and Casie's support the theory of response shift bias. Based on their comments about other Foundations students, the pre-post change scores captured true change. Pre-post and then-post pairwise comparisons of college belongingness scores were not significant. Even though my interviewees' responses make the college belongingness pre- and then-surveys appear to be interchangeable, I would caution against this assumption based solely on the lack of significant pairwise comparisons because then-survey scores were lower than post-survey scores and post-survey scores were lower than pre-survey scores.

Math mindset, math self-efficacy, and math belongingness were not significant (even with a less restrictive alpha), suggesting these dm-noncognitive factors as measured by this particular instrument may not be as susceptible to response shift. My interviews agree with this suggestion for math mindset and math belongingness. The then-survey for math self-efficacy better represents Robert and slightly better represents Casie, but the pre-survey is a slightly better representation of Mia. Because there are only slight differences for Mia and Casie, the pre- and then-surveys may be interchangeable, making the then-survey a potentially superior measure of math self-efficacy.

In very speculative summary of the above, pre- and then-surveys of math mindset, math self-efficacy, and math belongingness may be interchangeable (though a small amount of evidence suggests then-surveys may be better suited for measuring math self-efficacy). Pre-surveys may be better suited for measuring math equanimity. I do not have sufficient information to make an informed hypothesis about whether the pre- or then-survey is a truer measure of college belongingness. My research contradicts the claim that then-surveys signal larger program effects than pre-surveys, but the stable then- to post-survey math equanimity scores indicate possible biases explained by the implicit theory of stability or the theory of impression management.

Question 2: Outcomes and Changes in DM-Noncognitive Factors

I hypothesized changes in Foundations students' dm-noncognitive factors would predict their math course grades, math final exam grades, and math course attendance. An increase in math self-efficacy had a positive effect on grades in the course and on the final exam. This aligns with the research hypothesis. The positive effect of increased math self-efficacy on grades is significant even after accounting for the (even larger) positive association between students' grades and students' pre-survey reports of math self-efficacy. This shows that regardless of students' initial math self-efficacy, an increase in this dm-noncognitive factor is still significantly predictive of higher grades. Causality is not established, so it is unclear if success increased self-efficacy, vice versa, or if there was a lurking variable (e.g., instructor).

An increase in math equanimity had a negative effect on attendance, and there was a negative association between initial equanimity and attendance. In other words, the more a students' anxiety decreased, the less he or she attended class. Low initial math equanimity is also associated with higher course grade, which seems to indicate that maybe a little anxiety is good. This agrees with an assumption of arousal theorists that being devoid of math anxiety is not optimal for performance (Ma, 1999).

A word of caution about the interpretation of the findings from Research Question 2: Analyzing this research question was a major learning point for me. I was really interested in the relationship between changes in students' dm-noncognitive factors and changes in students' grades. I could not give students a pre-assessment because it would violate the course setup, and without a pre-assessment, I cannot discuss changes in their grades. When I posed this research question and formulated my hypothesis, it made intuitive sense—Of course an increase in dm-noncognitive factors would be associated with positive outcomes! Without any type of pre-assessment I certainly have intriguing findings, but I do not believe there is a straightforward interpretation of the findings. Furthermore, changes in–dm-noncognitive factors may not be a consistent predictor of semester outcomes because a change score on a dm-noncognitive factor scale should be considered relative to the initial score. For example, an increase of one point on a dm-noncognitive factor should be expected to have a different impact on semester outcomes for a student with a low initial score versus a student with a high initial score. I attempted to account for this by including pre-survey dm-noncognitive factor scores in my model in addition to change scores, but this model still assumes a universal effect of changes regardless of initial self-report.

Question 3: Evidence of Place Value Concept Transfer

I hypothesized Foundations students would demonstrate their ability to transfer their knowledge to non-traditional place value problems. A small number of students exhibited accuracy on the online DAPVU. Not surprisingly, students were most accurate on a familiar-systematic problem situation (Maria's Error Pattern). While no students correctly solved the unfamiliar-systematic problem situation (Chocolate Factory), 10% were accurate on the unfamiliar-nonsystematic problem situation (Rugolia), and 10% were accurate on the familiar-nonsystematic problem situation. My interviewees did not have accurate responses on the online DAPVU (with the exception of Robert being partially accurate on PS5). However, they mostly exhibited accurate or partially accurate responses in person. This was untrue of the problem situation similar to PS3 (Chocolate Factory). Mia attempted, but did not accurately solve, TA Chocolate Factory; neither Casie nor Robert attempted TA Chocolate Factory. Mia also attempted, but did not accurately solve, TA Space Shuttle.

Many students used accurate place value language to varying degrees of specificity while describing Bobby's understanding on PS2 (32%) and while describing Maria's error pattern on PS5 (48%). Students who used accurate place value language strongly favored using it to describe concepts on PS2 (75%) but behaviors on PS5 (69%). In both of these familiar-systematic problem situations, respondents are asked explicitly to evaluate the hypothetical student's understanding. I believe the discrepancy in type of language use relates to the problem setup: Bobby is not using an algorithm, but Maria is. For PS5, the respondent must carefully consider the procedure Maria is using, replicate the procedure, and then think about and verbalize why Maria has taken the steps she has taken; whereas, in PS2, the respondent is only asked to evaluate Bobby's understanding.

My interviewees used more place value language (or more sophisticated place value language) during the interview than on their online DAPVU. Robert used place value language to describe concepts online and during the interview, but Maria and Casie chiefly focused on behaviors when using place value language.

The majority of students did not provide much depth of analysis on PS2 and PS5; approximately 80-85% focused solely on behaviors. Approximately 60% did not recognize the mixed-grouping place value structure on PS1 and half of the remaining students had a conceptual error. None of the 16 students who were assigned a depth of understanding score on PS3 exhibited more than procedural understanding, and 11 of these students showed no evidence they recognized the base-four place value structure. Six out of the 10 students who were assigned a depth score on PS4 recognized the mixed-grouping place value structure but had computational or conceptual errors; three students correctly applied regrouping strategies, but some of these students left the computation incomplete or used unnecessary (accurate) regrouping strategies. Over all the online DAPVU problem situations, students exhibited mostly algorithmic understanding; students approached a tacit level of understanding on the familiar non-systematic problem situation (PS1).

In cases we were able to discuss the familiar-systematic problem situations (PS2, PS5 and TA Mark's Addition), my interviewees offered either the same or greater depth of analysis in person than they delivered online. The differences between online and in-person responses were most salient for Robert on the familiar-nonsystematic problem situations (PS1 and TA Space Shuttle); Robert moved from not showing evidence of understanding to receiving a top score in understanding, Casie's scores improved by one point, and Mia's scores remained the same. My interviewees evidenced better understanding on the unfamiliar-systematic problem situations (PS3 and TA Chocolate Factory) during the interview than they had online, but these improvements were minimal. Robert showed no expansions in his depth of understanding on the unfamiliar-nonsystematic problem situations (PS4 and TA Rugolia). Mia's score for depth of understanding on PS4 was a 1, but during the interview she evidence the deepest level of understanding.

As can be expected, I received much more information about the interviewees' ability to transfer their knowledge during the interviews than I did from their online responses. The interviewees exhibited algorithmic understanding on the online DAPVU, but they showed evidence of tacit understanding on more than one problem situation in person. This distinction may be attributed to the differences between written assessments and clinical interviews or it could be they had greater understanding of the types of problem situations as a result of prior exposure (i.e., learning from the test). Robert provided some evidence of explicit understanding. Mia and Casie did not provide much evidence of explicit understanding because they did not clearly articulate and convincingly justify concepts. For example, Casie was able to explain Mark's error pattern on TA Mark's Addition, but she was not able to explain why his procedure was invalid.

Up to this point, I have focused on what students transferred out and using the DAPVU as an SPS measure of transfer. Focusing attention on whether or not students "know that" or "know how," I found evidence of failed transfer (where they did not use applicable knowledge) and negative transfer (e.g., applying base-ten strategies to problem situations using a base other than ten). In some cases where it appears as though DAPVU students failed to transfer in their place value knowledge, it was clear they did not fully read the problem situation. Hence, these should not be deemed instances of failed transfer.

For some problem situations, I included prompts similar to the eagle challenge (See Schwartz et al. (2005)) to ascertain what students transferred in when they were unable to solve the problem situations. Some students said visual aids or watching someone else solve a similar problem would help them, but they did not mention their ideas about the mathematical content of the problem situations. Forty percent of students recognized the mixed-grouping place value structure on PS1, but were unsure how to perform calculations and predominately operated in base-ten. This shows they transferred in their understanding of base-ten. They noticed problem features, recognizing them as different from previous situations. For example, they realized there were differences between areas where they should compute in base-sixty and compute in base-ten on PS1. However, many were not sure how to compute in base-sixty. Several students resisted premature assimilation. They provided numeric responses, but described ways in which they were confused and questioned the accuracy of their computations. As mentioned before, the recognition that they may not have sufficient understanding is a prerequisite for cognitive incongruity. For example, Robert experienced cognitive incongruity (specifically, discoordination) on TA Space Shuttle. While monitoring his comprehension, he noticed the hundredths of a second column should be treated differently than the seconds column, but he was not sure how to interpret the hundredths of a second column. He persisted and eventually deduced 100 hundredths of a second is equivalent to 1 second, and he used this information to correctly solve the problem situation. Even if he had not made this deduction, his noticing of the problem features suggested he was prepared to learn how to solve the problem situation. Mia and Casie also evidenced their preparation for future learning by verbalizing some of what they transferred in to problem situations.

ADDRESSING COMPLICATIONS AND LIMITATIONS

This is an exploratory study, meant to inform an underrepresented area of math education research, in which I provide only speculative conclusions. Ideally, the pre-survey should be administered at the beginning of the course, but it was administered in either the second or third week of the course, after students potentially underwent changes in their dm-noncognitive factors. This muddies results for comparisons between the pre- and post-survey as well as comparisons between the pre- and then-survey. The post-survey was administered close to the end of the semester, perhaps too close to the administration of the final exam, and end-of-semester pressures may have negatively affected students' responses. If I were to redo this study, I would make every effort to administer the pre-survey on the first or second day of the course and make sure the post-survey is administered at least two weeks prior to the final exam.

Students were instructed not to begin the then-survey until they completed the post-survey, but there was no way to enforce these instructions or determine whether or not these instructions had been followed. If they did not follow instructions, the post-then-survey would behave as a less reliable, post&then-survey (or a perceived change measure). This issue could be resolved if instructors were willing to provide additional support (by handing out then-surveys only after students returned post-surveys) or if the researchers oversaw administration. The reasons behind the math equanimity and college belongingness response shift (Research Question 1B) are unclear. Including a control group, a measure of nonsubjective change, or a memory thentest could clarify findings. Also, a greater number of interviews could point to the reasons behind the response shift.

Research Question 2 was challenging to interpret. I did not have access to a pre-assessment with which I could have compared changes in dm-noncognitive factors to changes in mathematical success. Given enough participants, Research Question 2 could be further explored by looking at the possible moderation of pre-survey scores on change scores (i.e., the interaction). Responses could be categorized by level and direction of change (i.e., one unit increase in equanimity vs. two unit increases in equanimity) and compared to outcomes.

My desire to make the DAPVU a PFL transfer assessment conflicted with my goal to align it with the APVU. I avoided making references to place value, but some embedded references to place value may be appropriate to help students make connections and show they are able to "know with" (i.e., evidence their interpretive knowing). The DAPVU needed to be administered online; this rendered the established rubrics unusable and relegated my analyses to being purely exploratory. The online DAPVU was not a mandatory problem set and incentives were provided for completion, not on achievement, so students may not have invested as much time and effort as they would have otherwise. This applies both to the calculations and the explanations. This made it problematic to assess some of the facets of understanding, and results from evidence markers may be lower than they should be. This is supported by the interviews in which students demonstrated greater scores on many evidence markers in person than they did online.

The results of this study are not generalizable because I did not have access to a control group, there were only two participating institutions (both with strong buy-in), and student participation was not mandatory. A larger, randomized controlled trial would better address my research questions, but the overwhelming implementation of NMP would make it difficult to secure control groups. I did not receive sufficient interviews to draw substantial conclusions. The interview length could be decreased, but this would result in a tradeoff between the amount of information gained from each interviewee and the number of interviewees. If the interview length were decreased, it may be advisable to focus on fewer talking points for each participant or randomly delegating talking points to different interviewees.

CONTRIBUTIONS

Despite its limitations, I believe this study makes meaningful contributions. Foundations aims to improve students' math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness, but students only reported significant increases in their math equanimity (and college belongingness if a less restrictive alpha is used). Perhaps this means that Foundations, and courses similar to Foundations, could do more to specifically target the seemingly unaffected dm-noncognitive factors that are such strong predictors of student success.

In addition to responding to the threat of response shift bias, I investigated retrospective measures to determine if they could adequately replace traditional measures for the purposes of saving time and effort costs. There is some indication then-surveys may be sufficient measures of math mindset, math self-efficacy, and math belongingness for students in Foundations courses. My study showed then- and post-survey comparisons do not always show bigger program effects than pre- and post-survey comparisons; I did not encounter this issue with any of the constructs used in this study. A related goal involved finding valid and reliable short measures of dm-noncognitive factors to replace long, redundant measures. I believe the surveys used in this study afford researchers the ability to quickly and accurately assess five dm-noncognitive factors. Administrators and instructors may also use the surveys as pulse checks throughout the semester.

The think-aloud confirmed existing research by demonstrating how SPS measures of transfer might not sufficiently capture a student's true mathematical understanding—the interviews often illuminated student understanding that was not apparent in the online DAPVU. While the DAPVU still requires some tweaking to better assess students' ability to transfer their place value knowledge, it is mostly parallel to the validated APVU and may be used in a variety of ways. It could be used in the classroom to promote productive struggle and provide students an introduction to place value concepts in base n (where n $\neq 10$) prior to formal introduction of traditional base n computation. Because it is much shorter than the APVU, students may be more inclined to persist. I also believe the think-aloud problems and DAPVU problems would be excellent for a clinical teaching interview. There were several instances when interviewees asked questions that, had I been conducting a clinical teaching interview, could have been answered in ways that extended their understanding and provided me with greater information about their preparation for future learning.

FUTURE DIRECTIONS

I have learned a great deal from this study and, as with any research endeavor, my study inspired possible directions for future research. I would like to further explore the possible interchangeability of math mindset, math self-efficacy, and math belongingness pre-surveys and then-surveys in both Foundations courses and similar developmental mathematics courses. Students exhibited response shift on math equanimity and college belongingness, indicating surveys measuring these latent constructs may be more prone to biases than the other dm-noncognitive factors in this study. I am interested in conducting a future study that includes measures to explain the reason for these shifts.

My question about the relationship between changes in dm-noncognitive factors and changes in student success was not included in this study due to lack of a pre-assessment; I intend to consider this relationship more fully in subsequent work, by either using other success measures (e.g., placement assessments) or conducting research in a similar developmental course which allows a pre-assessment. This would provide information regarding which noncognitive factors are most important for impacting student willingness to take on challenges and productively persist in those challenges.

Finally, I aim to continue tackling difficult questions about transfer by modifying the DAPVU with information I gained from the evidence markers and the interviews. The modifications will likely further remove it from being parallel to the APVU because I will focus on trying to uncover students' interpretive knowing, so I will need to validate it through pilot testing and interviews.

OVERALL SUMMARY AND RELEVANCE

My dissertation study was principally exploratory and had multiple limitations. However, I was able to shed light on my research questions and arrived at interesting, and sometimes unexpected, conclusions. My primary goals involved exploring Foundations students' math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness, and how these dm-noncognitive factors relate to student outcomes and their ability to transfer their place value knowledge to novel situations. Secondarily, I investigated the utility and accuracy of practical measures meant to detect these dm-noncognitive factors.

The more math self-efficacious Foundations students became during the semester, the higher their course grades and final exam grades were. Lower initial math equanimity was also associated with better course grades. Foundations students' enhanced their math equanimity, but this enhanced equanimity was associated with a decrease in attendance. Math equanimity and college belongingness appear to be more susceptible to response shift than math mindset, math self-efficacy, and math belongingness. True pre-surveys may be superior indicators of initial math equanimity and then-surveys may potentially replace pre-surveys as initial indicators of math mindset, math self-efficacy, and math belongingness. Contrary to what much research has shown, comparisons between then- and post-surveys did not show larger program effects than comparisons between pre- and post-surveys.

My interviewees mostly demonstrated positive changes in their dm-noncognitive factors, though their surveys sometimes suggested otherwise. Foundations appeared to have a strong impact on two of these three students, and a moderate impact on one of the students, suggesting the surveys may not have detected change that occurred. Though the interviewees are not representative of all Foundations students, some of their comments about the other students in the course imply they were not alone in these improvements; and the interviewees attributed many of these changes to the presentation of the content.

Foundations explicitly or implicitly addresses the five dm-noncognitive factors used in my study, strategically stimulates productive persistence, and promotes transferability. These embedded interventions, on the face, look like they have a powerful impact on students. According to anecdotal accounts from faculty and my interviewees, Foundations students seem changed at the end of the semester—they appear more motivated to learn, more willing to wrestle with problems until they find a solution, and more confident sharing their ideas. Teaching Foundations requires strong instructor buy-in and has encouraged faculty to deeply reflect on their practice as educators. The curricular materials are quite dense, and the instructor must truly immerse him or herself in order to do the course justice.

I found mostly examples of failed transfer and negative transfer when treating the DAPVU as an SPS measure of transfer. I found minimal evidence of what students transferred in, but this could be due to the online structure and lack of participant engagement. The findings do not indicate Foundations students are unable to transfer their knowledge to novel place value problems—the findings merely show I was unable to adequately detect their ability to transfer out and whether or not they were prepared to transfer their learning in to the assessment or future learning experiences. The interviewees evidenced more accuracy, use of descriptive language, and depth of analysis and understanding during the interviews, and they exhibited greater evidence of what they transferred in to the interviews than what they transferred in to the online DAPVU.

Extensive research has shown the importance of noncognitive factors for K-12 students and university students, but developmental mathematics students are a historically neglected demographic in relation to these types of studies. Researchers have recently begun to focus more attention on studying the impact of psychosocial interventions on developmental mathematics students' noncognitive factors and mathematics achievement. This is, in part, due to greater recognition of the influence of noncognitive factors on all students. It also stems from policy shifts aimed at increasing community college students' persistence and chances of receiving a post-secondary degree. Despite recent developments, there exists a need for more research regarding what factors embolden developmental mathematics students to take on challenges and persist when confronted with failure, and what helps them apply what they learn in unusual situations. My dissertation research provided some insight into these greatly needed areas of research and my future research will continue to focus on finding additional ways to help fill this need.

Appendix A: IRB Approval

OFFICE OF RESEARCH SUPPORT

THE UNIVERSITY OF TEXAS AT AUSTIN

P.O. Box 7426, Austin, Texas 78713 · Mail Code A3200
(512) 471-8871 · FAX (512) 471-8873

FWA # 00002030

Date: 01/02/15

PI: Philip U Treisman

Dept: Science, Technology, Engineering and Mathematics

Title: Post Hoc Discernment of Pro-Mathematical Behavior and Concept Transfer

Re: IRB Exempt Determination for Protocol Number 2014-10-0079

Dear Philip U Treisman:

Recognition of Exempt status based on 45 CFR 46.101(b)(2).

Qualifying Period: 01/02/2015 to 01/01/2018. *Expires 12 a.m. [midnight] of this date.*
A continuing review report must be submitted in three years if the research is ongoing.

Responsibilities of the Principal Investigator:

Research that is determined to be Exempt from Institutional Review Board (IRB) review is not exempt from ensuring protection of human subjects. The Principal Investigator (PI) is responsible for the following throughout the conduct of the research study:

1. Assuring that all investigators and co-principal investigators are trained in the ethical principles, relevant federal regulations, and institutional policies governing human subject research.

2. Disclosing to the subjects that the activities involve research and that participation is voluntary during the informed consent process.

3. Providing subjects with pertinent information (e.g., risks and benefits, contact information for investigators and ORS) and ensuring that human subjects will voluntarily consent to participate in the research when appropriate (e.g., surveys, interviews).

4. Assuring the subjects will be selected equitably, so that the risks and benefits of the research are justly distributed.

5. Assuring that the IRB will be immediately informed of any information or unanticipated problems that may increase the risk to the subjects and cause the category of review to be reclassified to expedited or full board review.

6. Assuring that the IRB will be immediately informed of any complaints from subjects regarding their risks and benefits.

7. Assuring that the privacy of the subjects and the confidentiality of the research data will be maintained appropriately to ensure minimal risks to subjects.

8. Reporting, by submission of an amendment request, any changes in the research study that alter the level of risk to subjects.

These criteria are specified in the PI Assurance Statement that was signed before determination of exempt status was granted. The PI's signature acknowledges that they understand and accept these conditions. Refer to the Office of Research Support (ORS) website www.utexas.edu/irb for specific information on training, voluntary informed consent, privacy, and how to notify the IRB of unanticipated problems.

1. Closure: Upon completion of the research study, a Closure Report must be submitted to the ORS.

2. Unanticipated Problems: Any unanticipated problems or complaints must be reported to the IRB/ORS immediately. Further information concerning unanticipated problems can be found in the IRB Policies and Procedure Manual.

3. Continuing Review: A Continuing Review Report must be submitted if the study will continue beyond the three year qualifying period.

4. Amendments: Modifications that affect the exempt category or the criteria for exempt determination must be submitted as an amendment. Investigators are strongly encouraged to contact the IRB Program Coordinator(s) to describe any changes prior to submitting an amendment. The IRB Program Coordinator(s) can help investigators determine if a formal amendment is necessary or if the modification does not require a formal amendment process.

If you have any questions contact the ORS by phone at (512) 471-8871 or via e-mail at orsc@uts.cc.utexas.edu.

Sincerely,

James P. Wilson

James Wilson, Ph.D.
Institutional Review Board Chair

Appendix B: Consent Description

Consent Description for Participation in Research

Title: Post hoc discernment of mathematical behavior and concept transfer

Introduction

The purpose of this form is to provide you information that may affect your decision as to whether or not to participate in this research study. This form is yours to keep. The person performing the research will answer any of your questions. Read the information below and ask any questions you might have before deciding whether or not to take part. If you decide to be involved in this study, you will sign a consent form.

Purpose of the Study

You have been asked to participate in a research study about math attitudes and application of math knowledge to new contexts. One purpose of this study is determine math attitudes of students in non-traditional developmental math classes, measure whether or not their attitudes change over the course of the semester, and measure whether or not those attitudes are related to how students perform on unfamiliar math tasks. Another purpose is to help researchers determine the pros and cons of using different types of surveys, and help researchers find practical methods for assessing student change in non-traditional developmental math courses.

What will you be asked to do?

If you agree to participate in this study,

At the beginning of the semester you will be asked to:

- Complete a short survey about your math attitudes (<u>in class</u>)

At the end of the semester you will be asked to:

- Complete a short survey about your math attitudes and background (<u>in class</u>)
- Complete an online math assessment
- A select number of participants may be asked to participate in interviews that will take less than one hour. If you participate in an interview, your participation will be audio recorded. You may agree to any other parts of the study without agreeing to participate in interviews.

If you agree to participate, you understand that additional information will be collected on you from the institution. This includes gender, age, ethnicity, final exam grade, academic course grade, and attendance records.

This study will take approximately 1 hour total (2 hours total if you do an interview), spread out over the course of the entire semester and will include approximately 200 study participants.

What are the risks involved in this study?

There are no foreseeable risks to participating in this study. Your course grade will not be affected in any way due to your decision to participate or not participate.

What are the possible benefits of this study?

Your participation can help inform: creators of developmental mathematics curriculum about positive and/or negative aspects of a curriculum, educational researchers about student's math attitudes and how students apply their math knowledge, and the design of surveys and other measures for research. You may be compensated for your participation.

Do you have to participate?

No, your participation is voluntary. You may decide not to participate at all or, if you start the study, you may withdraw at any time. Withdrawal or refusing to participate will not affect your relationship with your college or The University of Texas at Austin in any way.

If you would like to participate in the study, on the first page of the survey, please sign your name and fill out your contact information. Your information will be used only for the purposes of this study and will not be shared. You will not receive spam e-mails.

Will there be any compensation?

- If you complete the <u>two in-class surveys</u> about math attitudes, beliefs, and background at the beginning and end of semester, you will be entered into a drawing. A randomly selected participant will receive an HEB gift card for $100.
- If you complete the <u>two in-class surveys</u> about math attitudes, beliefs, and background at the beginning and end of semester AND the <u>online math assessment</u> at the end of the semester, you will automatically receive an HEB gift card for $15 and be entered into a drawing. A randomly selected participant will receive an HEB gift card for $100.
- If you complete the <u>two in-class surveys</u> about math attitudes, beliefs, and background at the beginning and end of semester AND the <u>online math assessment</u> at the end of the semester AND participate in an audio-recorded <u>interview</u>, you will automatically receive an HEB gift card for $25 and be entered into a drawing. A randomly selected participant will receive an HEB gift card for $100.

How will your privacy and confidentiality be protected if you participate in this research study?

The records of this study will be stored securely and kept confidential. Your privacy and the confidentiality of your data will be protected through the following methods:

- Only the researchers will know whether or not you have consented to participate; your course instructor will not be told whether or not you have agreed to participate.
- All data will be labeled with a randomly assigned number in place of any personally identifying information. Once the data has been analyzed the master key connecting you to your study ID will be destroyed to protect your identity.

- Consent forms will be locked in a secure cabinet, accessible only by the principal investigator. Paper data will be locked in a separate secure cabinet, accessible only by the principal investigator. Data kept in electronic files will be secured through security access codes. All online data will be collected through a secure system.
- You may participate in this study without participating in an interview. If you choose to participate in an interview for this study, you will be audio recorded. Any audio recordings will be stored securely and only the research team will have access to the recordings. The digital recordings will be coded so that no personally identifying information is visible on them. Recordings will be kept until they have been transcribed and then the will be erased. Transcriptions of the recordings will be stored in an electronic file with no identifying information except your study ID and will be secured through security access codes.
- All publications will exclude any information that will make it possible to identify you as a participant.
- Throughout the study, researchers will notify you of new information that may become available and that might affect your decision to remain in the study.

If it becomes necessary for the Institutional Review Board to review the study records, information that can be linked to you will be protected to the extent permitted by law. Your research records will not be released without your consent unless required by law or a court order. The data resulting from your participation may be made available to other researchers in the future for research purposes not detailed within this consent form. In these cases, the data will contain no identifying information that could associate it with you, or with your participation in any study.

Whom to contact with questions about the study?

Prior, during or after your participation you can contact the researcher **Stephanie Baker Peacock** by sending an email to **speacock@math.utexas.edu** for any questions or if you feel that you have been harmed.

Whom to contact with questions concerning your rights as a research participant?

For questions about your rights or any dissatisfaction with any part of this study, you can contact, anonymously if you wish, the Institutional Review Board by phone at (512) 471-8871 or email at orsc@uts.cc.utexas.edu.

Appendix C: Consent Form

If you agree to participate in this study, please acknowledge the following statement by signing and providing your contact information. Please print legibly.

I have a copy of the consent form with the study's purpose, procedures, possible benefits and risks. I understand that I can contact the researcher at any time before or during the study to ask questions or ask to be removed from the study without it affecting my standing with my college. I understand that my participation is voluntary and confidential and my responses will not be shared in a way that identifies me as a participant. I understand that, by signing this form, I am not waiving any legal rights.

Sign name to agree		**Date:**
Email address (print)		
First name (print)		
Last name (print)		

Do you wish to participate in an interview about math beliefs and concepts? You may participate in the study without participating in an interview and you may change your mind at any time. If you agree to participate in an audio-recorded interview, please sign and date below.

Sign name to agree to an interview		**Date:**

Appendix D: Pre-Survey Directions for Instructors

SURVEY 1

Dear [Instructor],

Thank you so much for your willingness to assist with this study. Your time is very valuable, so I have tried to make this as unobtrusive as possible. If you have any questions or concerns, please contact me at [XXX-XXX-XXXX] or speacock@math.utexas.edu. I have outlined the actions asked of you below.

Sincerely,
Stephanie Baker Peacock

SURVEY 1 PACKET CONTENTS

- This information sheet
- Consent Description for the students to keep. There are extras in case students lose theirs.
- Survey 1 (3-5 minutes in class at the beginning of semester)
- Scantrons for Survey 1

AS CLOSE TO THE BEGINNING OF THE SEMESTER AS POSSIBLE
Please write the following on the board:

Identification: Student ID
Special Codes: [*Please write your Teacher ID*]

Please pass out one Consent Description, one survey, and one scantron per student.
Please read the following statement (or rephrase as you see fit):

A student, named Stephanie Peacock, at The University of Texas at Austin has asked us to help with a study she is conducting to learn about student math attitudes and how students think about different types of math problems. She is conducting the study as part of a requirement to obtain her doctoral degree and is very interested in your ideas about math. She provided me with this description to read to you because she is unavailable to meet with you personally at the beginning of the semester. There are more details in the Consent Description I just gave you.

Most importantly, she wants you to know that your participation is completely voluntary, you can stop participating at any time, you can contact her if

you have questions, your responses will have no impact on your grade, and she is the only person that will know your responses and whether or not you are participating. I also will not know if you are participating.

There are benefits for participation. You will be given a chance to really reflect on your ideas about math and you could receive HEB gift cards. There are 2 surveys you can do in class, a few math problems you can do online, and an interview you may do in person with Stephanie. As you see on page 2, **if you complete the 2 surveys in class you will be entered into a drawing and a randomly selected person will receive an HEB gift card for $100. You will do the first survey now and it will take less than 5 minutes**. The survey at the end of the semester will take less than 10 minutes.

If you do the 2 surveys and also do the online math problems, you will get a gift card for $15 and be entered into the drawing for $100.

If you do the 2 surveys, the online math problems, and an audio-recorded interview, you will get a $25 gift card and be entered into the drawing for $100.

The most you could receive for participating is $125. Stephanie would use your course grade and final exam grade to see if there is a relationship between grades and beliefs. Stephanie needs this information to do the study, but your personal information would not be shared with anyone except her.

This is, in no way, a judgment of how good you are at math. There are no right or wrong answers when it comes to your beliefs. She just wants to try and understand how you think about math. If you want to participate, please fill out the information on the first page of the survey I gave you. **Please write legibly because she will need to contact you with a link for the math problems and to give you gift cards. If you also want to do an interview, sign and date the bottom box.** Then fill in your first and last name, date, and course in the places provided on the scantron. Bubble in your student ID in the box labeled "Identification." Please **bubble** my **Teacher ID** in the box labeled "Special Codes." My Teacher ID is on the board. If you do not know your student ID number, please bubble in your first and last name in the spaces provided. Then complete the surevey on the scantron.

Please try to answer as truthfully as possible. If you decide to participate, put your survey pages together with your scantron in this envelope [*Survey 1 envelope*]. Please keep your survey together with your scantron when you put it in the envelope. If you don't want to participate you can also put your scantron with your survey in the envelope. This way, I won't know who is participating. You can all keep the 3-page Consent Description in case you want to contact Stephanie. Also, she wants me to thank you for considering participating because your responses are invaluable to her work. [*Please give* *the students 5-8 minutes to complete the consent and survey. Please try to minimize your chance of seeing who is participating*.]

Thank you so much for your assistance. When students have completed this survey, please ***ask them to put their surveys with their scantrons into the envelope*** *marked Survey #1. Please seal the envelope once all of the students have done this and give the envelope to [campus contact].*

Appendix E: Pre-Survey

You will be providing answers to all of the following questions **on your scantron**. For each question, please completely bubble in **only one** response. There are no right or wrong responses, only the way you feel about each statement or question.

On your scantrons, please write your **name**, **date** and **course** in the **spaces provided**. Please **bubble** your **student ID** in the box labeled “Identification.” Please **bubble** your **Teacher ID** in the box labeled “Special Codes.”

*If you do not know your student ID number, please bubble in your first and last name in the spaces provided. Your Teacher will provide you with his/her Teacher ID.

The 4 items below refer to things that may cause fear or tension. Don't think about it too much; just mark the response that first comes to mind.

1. **How anxious would you feel listening to a lecture in math class?**
   a) Extremely anxious b) Very anxious c) Moderately anxious d) Slightly anxious e) Not at all anxious

2. **How anxious would you feel taking a math test?**
   a) Extremely anxious b) Very anxious c) Moderately anxious d) Slightly anxious e) Not at all anxious

3. **How anxious would you feel signing up for a course in math?**
   a) Extremely anxious b) Very anxious c) Moderately anxious d) Slightly anxious e) Not at all anxious

4. **How anxious would you feel the moment before you got a math test back?**
   a) Extremely anxious b) Very anxious c) Moderately anxious d) Slightly anxious e) Not at all anxious

Read the 3 statements below and mark how much you agree or disagree with each statement.

5. **You have a certain amount of math intelligence and you really can't do much to change it.**
   a) Strongly agree b) Agree c) Mostly agree d) Mostly disagree e) Disagree f) Strongly disagree

6. **Your math intelligence is something about you that you can't change very much.**
   a) Strongly agree b) Agree c) Mostly agree d) Mostly disagree e) Disagree f) Strongly disagree

7. **You can learn new things, but you can't really change your basic math intelligence.**
   a) Strongly agree b) Agree c) Mostly agree d) Mostly disagree e) Disagree f) Strongly disagree

The next 5 items are about you as a student in your current math class. How much is each of the following statements true or not true of you? There are no right or wrong answers, so please mark the response that best describes what you think.

8. I'm certain I can master the skills taught in my math class this semester.
a) Not at all true of me b) Slightly true of me c) About halfway true of me d) Mostly true of me e) Very true of me

9. I'm certain I can figure out how to do the most difficult work in my math class.
a) Not at all true of me b) Slightly true of me c) About halfway true of me d) Mostly true of me e) Very true of me

10. I can do almost all the work in my math class if I don't give up.
a) Not at all true of me b) Slightly true of me c) About halfway true of me d) Mostly true of me e) Very true of me

11. Even if the work in my math class is hard, I can learn it.
a) Not at all true of me b) Slightly true of me c) About halfway true of me d) Mostly true of me e) Very true of me

12. I can do even the hardest work in my math class if I try.
a) Not at all true of me b) Slightly true of me c) About halfway true of me d) Mostly true of me e) Very true of me

The 2 items below refer to your thoughts about your current math class or college. Mark how often, if ever, you wonder these things.

13. When you think about your math class, how often, if ever, do you wonder: *Maybe I don't belong here.*
a) Always b) Frequently c) Sometimes d) Hardly ever e) Never

14. When you think about your college, how often, if ever, do you wonder: *Maybe I don't belong here.*
a) Always b) Frequently c) Sometimes d) Hardly ever e) Never

Thank you for taking the time to complete this survey. You will be given a follow-up survey near the end of the semester. If you have any additional comments, please use the space below or on the back of this page.

Appendix F: Post-Survey Directions for Instructors

SURVEY 2

Dear [Instructor],

Thank you so much for your willingness to assist with this study. Your time is very valuable, so I have tried to make this as unobtrusive as possible. **Please give students 10 minutes to complete the survey** and try minimizing your chance of seeing who is participating. When students have completed this survey, please **ask them to put their surveys with their scantrons into the envelope marked Survey #2.** Please seal the envelope once all of the students have done this. Please return these surveys to [campus contact] by Monday, November 16. If you have any questions or concerns, please contact Stephanie: XXX-XXX-XXX; speacock@math.utexas.edu. I truly appreciate your assistance.

Sincerely,
Stephanie Baker Peacock

PACKET CONTENTS

- This information sheet
- Example scantron
- Scantrons for students
- Survey 2 for students
- Consent descriptions
- Survey 2 envelope (Students will put their surveys and scantrons in here)

Please write the following on the board:

Special Codes:

Please pass out one survey and one scantron per student. Please give consent descriptions to any students who want them. Please read the following statement (or rephrase as you see fit):

At the beginning of the semester, you were asked to participate in a study that is being conducted by Stephanie Peacock, a student at The University of Texas at Austin, so she may complete her degree. The study is about how students think about math and different types of math problems. The first survey had 14 questions and this survey has 37 questions. You should have received a consent description at the beginning of the semester, but I have extras here if you would like to look at it again or if you want a copy. **Participation is voluntary and your responses are confidential.**

Closer to the end of the semester, Stephanie will give you the opportunity to do some math problems. She will also do some interviews with people who are interested. Whether or not you do the math problems and interviews **you will be entered into a drawing for a $100 HEB gift card for completing the surveys**. **If you also do the math problems, you will automatically receive a $15 HEB gift card. If you also do an interview, you will automatically get a $25 HEB gift card.**

This is, in no way, a judgment of how good you are at math. There are no right or wrong answers when it comes to your beliefs. Stephanie just wants to try and understand how you think about math. So, please try to answer as truthfully as possible.

CONSENT FORM ON FIRST PAGE OF SURVEY

The first page of the survey has a consent form. **You may have already filled out a consent form, but you should do it again just in case she wasn't able to read your writing. If she can't match your surveys to your consent or if you don't fill in your email address, she won't have any way to pay you. If you also want to do an interview, sign and date the bottom box.**

SCANTRON

You must use pencil on the scantron and try to avoid making stray marks. Page 2 tells you what you should complete on your scantron. Please fill in your name. If you don't fill in your name, she will not be able to match your scantron to your consent and won't be able to pay you. Fill in your ID number, birthdate, and sex. Fill the first four columns of "special codes" with the number I have written on the board. On the upper right of the scantron, sign your name, write my name, and write today's date. Bubble in one response for each question on numbers 1 through 37.

SURVEY FORMAT

This survey is probably different from ones you have seen before. There are three sections. **Please complete the survey in order; don't move on until you have finished each section and don't go back to a previous section once you have completed it.** The questions in the first and second section will look very similar, so please pay attention to the directions when responding to those questions.

When you have completed the consent form and survey, please put your survey pages together with your scantron in this envelope [*Survey 2 envelope*]. Stephanie wanted me to thank you in advance for thinking about participating because your responses will be very helpful for her research.

Thanks again!

Appendix G: Post-Then-Survey

PLEASE USE A PENCIL TO FILL IN THE FOLLOWING INFORMATION ON YOUR SCANTRON:

- Last Name, First Name, Middle Initial – Write and bubble your name
- Identification – Write and bubble your Student ID if you know it
- Birthdate – Write and bubble your birthdate
- Sex – Bubble Male or Female
- Special Codes – Write and bubble the code given to you by your instructor (in the A, B, C, D columns)
  Signature – Sign your name
- Instructor – Write your instructor's name
- Date – Write today's date
- Questions 1-37 – Bubble in completely and avoid stray marks

*All of your information is confidential, but the researcher needs it in order to match your consent to your survey and to distribute gift cards.

*You will be providing answers to all of the following questions **on your scantron**. Please write in pencil. For each question, please completely bubble in **only one** response. <u>There are no right or wrong responses, only the way you feel about each statement or question</u>.

Mathematical Behaviors Survey—End of Semester

The 4 items below refer to things that may cause fear or tension. Don't think about it too much; just mark the response that first comes to mind.

1. **How anxious would you feel listening to a lecture in math class?**
   a) Extremely anxious b) Very anxious c) Moderately anxious d) Slightly anxious e) Not at all anxious

2. **How anxious would you feel taking a math test?**
   a) Extremely anxious b) Very anxious c) Moderately anxious d) Slightly anxious e) Not at all anxious

3. **How anxious would you feel signing up for a course in math?**
   a) Extremely anxious b) Very anxious c) Moderately anxious d) Slightly anxious e) Not at all anxious

4. **How anxious would you feel the moment before you got a math test back?**
   a) Extremely anxious b) Very anxious c) Moderately anxious d) Slightly anxious e) Not at all anxious

Read the 3 statements below and mark how much you agree or disagree with each statement.

5. **You have a certain amount of math intelligence and you really can't do much to change it.**
   a) Strongly agree b) Agree c) Mostly agree d) Mostly disagree e) Disagree f) Strongly disagree

6. **Your math intelligence is something about you that you can't change very much.**
   a) Strongly agree b) Agree c) Mostly agree d) Mostly disagree e) Disagree f) Strongly disagree

7. **You can learn new things, but you can't really change your basic math intelligence.**
   a) Strongly agree b) Agree c) Mostly agree d) Mostly disagree e) Disagree f) Strongly disagree

Mathematical Behaviors Survey—End of Semester

The next 5 items are about you as a student in your current math class. How much is each of the following statements true or not true of you? There are no right or wrong answers, so please mark the response that best describes what you think.

8. I'm certain I can master the skills taught in my math class this semester.
a) Not at all true of me b) Slightly true of me c) About halfway true of me d) Mostly true of me e) Very true of me

9. I'm certain I can figure out how to do the most difficult work in my math class.
a) Not at all true of me b) Slightly true of me c) About halfway true of me d) Mostly true of me e) Very true of me

10. I can do almost all the work in my math class if I don't give up.
a) Not at all true of me b) Slightly true of me c) About halfway true of me d) Mostly true of me e) Very true of me

11. Even if the work in my math class is hard, I can learn it.
a) Not at all true of me b) Slightly true of me c) About halfway true of me d) Mostly true of me e) Very true of me

12. I can do even the hardest work in my math class if I try.
a) Not at all true of me b) Slightly true of me c) About halfway true of me d) Mostly true of me e) Very true of me

The 2 items below refer to your thoughts about your current math class or college. Mark how often, if ever, you wonder these things.

13. When you think about your math class, how often, if ever, do you wonder: *Maybe I don't belong here.*
a) Always b) Frequently c) Sometimes d) Hardly ever e) Never

14. When you think about your college, how often, if ever, do you wonder: *Maybe I don't belong here.*
a) Always b) Frequently c) Sometimes d) Hardly ever e) Never

*****PLEASE DO NOT START THIS SECTION UNTIL YOU HAVE COMPLETED THE PREVIOUS SECTIONS*****

For the following questions, please evaluate each statement according to **how well it best described you at the beginning of this semester.** This survey gives you the chance to retrospectively assess your beginning of semester behavior, using the information you gained during the course of the semester. Think back to when you began this semester. Now that you have been in your math class for a while, how would you rate yourself as having been before?

You may remember how you rated yourself on these items when you first took this survey at the beginning of the semester. Please do not simply recall your original ratings. These ratings are to reflect your current opinion of your pre-semester behavior, based on the knowledge, ability, or awareness you gained during the course of the semester. Do not worry whether these ratings agree or disagree with your earlier ratings. Blacken the circles that most clearly described you as you were at the beginning of this semester.

Mathematical Behaviors Survey—Think back to the beginning of the semester

For the following questions, please bubble the one response that would have **best described you at the beginning of this semester.** There are no right or wrong responses, only the way you feel about each statement or question.

The next 5 items are about you as a student in your math class at the beginning of the semester. How much was each of the following statements true or not true of you? There are no right or wrong answers, so please mark the response that best describes what you think.

15. I'm certain I can master the skills taught in my math class this semester.
a) Not at all true of me b) Slightly true of me c) About halfway true of me d) Mostly true of me e) Very true of me

16. I'm certain I can figure out how to do the most difficult work in my math class.
a) Not at all true of me b) Slightly true of me c) About halfway true of me d) Mostly true of me e) Very true of me

17. I can do almost all the work in my math class if I don't give up.
a) Not at all true of me b) Slightly true of me c) About halfway true of me d) Mostly true of me e) Very true of me

18. Even if the work in my math class is hard, I can learn it.
a) Not at all true of me b) Slightly true of me c) About halfway true of me d) Mostly true of me e) Very true of me

19. I can do even the hardest work in my math class if I try.
a) Not at all true of me b) Slightly true of me c) About halfway true of me d) Mostly true of me e) Very true of me

Mathematical Behaviors Survey—Think back to the beginning of the semester

Read the 3 statements below and mark how much you would have agreed or disagreed with each statement at the beginning of the semester.

20. You have a certain amount of math intelligence and you really can't do much to change it.

a) Strongly agree b) Agree c) Mostly agree d) Mostly disagree e) Disagree f) Strongly disagree

21. Your math intelligence is something about you that you can't change very much.

a) Strongly agree b) Agree c) Mostly agree d) Mostly disagree e) Disagree f) Strongly disagree

22. You can learn new things, but you can't really change your basic math intelligence.

a) Strongly agree b) Agree c) Mostly agree d) Mostly disagree e) Disagree f) Strongly disagree

The 2 items below refer to your thoughts about your current math class or college. Mark how often, if ever, you would wonder these things at the beginning of the semester.

23. When you think about your math class, how often, if ever, do you wonder: *Maybe I don't belong here.*

a) Always b) Frequently c) Sometimes d) Hardly ever e) Never

24. When you think about your college, how often, if ever, do you wonder: *Maybe I don't belong here.*

a) Always b) Frequently c) Sometimes d) Hardly ever e) Never

Mathematical Behaviors Survey—Think back to the beginning of the semester

The 4 items below refer to things that may cause fear or tension. Mark the response that most closely matches how you felt at the beginning of the semester.

25. How anxious would you feel listening to a lecture in math class?
a) Extremely anxious b) Very anxious c) Moderately anxious d) Slightly anxious e) Not at all anxious

26. How anxious would you feel taking a math test?
a) Extremely anxious b) Very anxious c) Moderately anxious d) Slightly anxious e) Not at all anxious

27. How anxious would you feel signing up for a course in math?
a) Extremely anxious b) Very anxious c) Moderately anxious d) Slightly anxious e) Not at all anxious

28. How anxious would you feel the moment before you got a math test back?
a) Extremely anxious b) Very anxious c) Moderately anxious d) Slightly anxious e) Not at all anxious

Please answer these 9 questions to the best of your ability. The data you share will not be used to personally identify you, and will not be passed on to anyone else. **For each question, please bubble in only one response.**

29. Gender
    a. Male
    b. Female

30. Age group
    a. Under 18
    b. 18 to 19
    c. 20 to 21
    d. 22 to 24
    e. 25 to 29
    f. 30 to 39
    g. 40 to 49
    h. 50 to 64
    i. 65+

31. Race/ethnicity
    a. American Indian or other Native American
    b. Asian, Asian American, or Pacific Islander
    c. Native Hawaiian
    d. Black or African American, Non-Hispanic
    e. White, Non-Hispanic
    f. Hispanic, Latino, or Spanish
    g. Other/Multi-racial

32. Enrollment status during Fall 2015
    a. Enrolled part time
    b. Enrolled full time

33. Job status during Fall 2015
    a. Not currently employed
    b. Part time: Fewer than 20 hours per week
    c. Part time: 20 to 25 hours per week
    d. Part time: 26 to 30 hours per week
    e. Part time: 31 to 35 hours per week
    f. Part time: 36 to 39 hours per week
    g. Full time: Work 40 or more hours per week

34. How many children/dependents do you have?
   a. None
   b. 1
   c. 2
   d. 3+

35. Is English your native (first) language?
   a. Yes
   b. No

36. What is the highest level of education obtained by your **mother**?
   a. Not a high school graduate
   b. High school diploma or GED
   c. Some college, did not complete degree
   d. Associate degree
   e. Bachelor's degree
   f. Master's degree/1st professional
   g. Doctoral degree
   h. Unknown or Other

37. Are you, or was one of your siblings, the first in your family to go to college?
   a. Yes
   b. No

Thank you for taking the time to complete this survey. If you have any additional comments, please use the space on the back of this page. ***Place these pages with your scantron in the folder provided.***

Appendix H: Pre-, Post-, and Then-Survey Response Rates

Tables 27 and 28 list the number of consented Foundations students who completed each of the pre-, post-, and then-surveys of dm-noncognitive factors in College A and College B, respectively. Each table lists the instructor ID and class section ID. The bottom row of each table lists the percentage response rate for each survey.

Table 27

Pre-, Post-, and Then-Survey Frequencies and Response Rates for Consented Foundations Students at College A

Class	*n*	Math Equanimity			Math Mindset			Math Self-Efficacy			Math Belongingness			College Belongingness		
		Pre	**Post**	**Then**	**Pre**	**Post**	**Then**	**Pre**	**Post**	**Then**	**Pre**	**Post**	**Then**	**Pre**	**Post**	**Then**
1-1	20	17	14	14	17	14	14	17	14	14	17	14	14	17	14	14
1-2	15	10	12	12	9	12	12	9	12	12	8	11	12	8	12	12
2-3	21	19	14	14	19	14	14	19	14	14	19	14	14	19	14	14
2-4	14	12	14	14	12	14	14	12	14	14	12	14	14	12	14	14
2-5	16	16	11	11	16	11	11	16	11	11	16	11	11	16	11	11
3-6	21	19	17	17	19	17	17	18	17	17	17	17	17	17	17	17
4-7	14	11	9	9	11	9	9	11	9	9	11	9	9	11	9	9
4-8	21	18	17	16	18	17	16	18	17	17	18	17	16	18	17	16
5-9	16	15	9	9	15	9	9	15	9	9	15	9	9	14	9	9
6-10	23	13	18	17	13	18	18	13	18	18	13	18	18	13	18	18
7-11	14	14	0	0	14	0	0	14	0	0	14	0	0	14	0	0
8-12	11	11	4	3	11	4	3	11	4	3	11	4	3	11	4	3
8-13	8	8	3	3	8	3	3	8	3	3	8	3	3	8	3	3
9-14	7	7	2	2	7	2	2	7	2	2	7	2	2	7	2	2
10-15	19	18	10	10	18	10	10	17	10	10	17	10	10	17	10	10
10-16	15	13	11	11	13	11	11	13	11	11	13	11	11	13	11	11
11-17	21	19	13	11	19	13	11	19	13	11	18	13	11	18	13	11
12-18	13	12	5	5	12	5	5	12	5	5	12	5	5	12	5	5
12-19	3	3	0	0	3	0	0	3	0	0	3	0	0	3	0	0
(*n* = 292)		**87%**	**63%**	**61%**	**87%**	**63%**	**61%**	**86%**	**63%**	**62%**	**85%**	**62%**	**61%**	**85%**	**63%**	**61%**

Note. The "Class" column represents the 12 instructors and 19 class sections; the format is Instructor ID-Section ID. Column "*n*" lists the number of consented students in each corresponding section. The last row lists the percentage response rates of consented students for each survey.

Table 28

Pre-, Post-, and Then-Survey Frequencies and Response Rates for Consented Foundations Students at College B

Class	***n***	**Math Equanimity**			**Math Mindset**			**Math Self-Efficacy**			**Math Belongingness**			**College Belongingness**		
		Pre	**Post**	**Then**	**Pre**	**Post**	**Then**	**Pre**	**Post**	**Then**	**Pre**	**Post**	**Then**	**Pre**	**Post**	**Then**
13-20	21	19	17	17	19	17	17	19	17	17	19	17	17	19	17	17
14-21	25	25	14	14	25	14	14	25	14	14	25	14	14	25	14	14
14-22	16	13	8	8	13	8	8	13	8	8	13	8	8	12	8	8
15-23	26	24	21	17	24	21	18	24	21	19	24	20	17	24	20	17
16-24	23	22	7	7	22	7	7	22	7	7	22	7	7	22	7	7
17-25	15	15	8	8	15	8	8	15	8	8	15	8	8	15	8	8
18-26	26	23	16	16	23	16	16	23	16	16	23	16	16	23	16	16
18-27	24	22	12	12	22	12	12	22	12	12	21	12	12	21	12	12
18-28	26	24	15	15	24	15	15	24	15	15	24	15	15	23	15	15
18-29	22	22	9	9	22	9	9	22	9	9	22	9	9	21	9	9
18-30	16	16	10	9	16	10	9	16	10	9	16	10	9	14	10	9
19-31	23	21	11	11	21	11	11	20	11	11	20	11	11	20	11	11
20-32	15	15	12	12	15	12	11	15	12	11	15	12	11	15	12	11
21-33	27	26	11	10	26	11	10	26	11	10	26	10	10	26	10	10
(*n* = 305)		**94%**	**56%**	**54%**	**94%**	**56%**	**54%**	**94%**	**56%**	**54%**	**93%**	**55%**	**54%**	**92%**	**55%**	**54%**

Note. The "Class" column represents the 9 instructors and 14 class sections; the format is Instructor ID-Section ID. Column "*n*" lists the number of consented students in each corresponding section. The last row lists the percentage response rates of consented students for each survey.

Appendix I: Participant Demographics

Participants reported demographic data on questions 29-37 on the end-of-semester survey. Tables 29, 30, and 31 contain the demographic characteristics of the sample. The variables that were treated as categorical in the analyses are in Table 29. The variables that were treated as continuous in the analyses are presented in categorical format in Table 30 and in continuous format in Table 31. Table 29 and Table 30 also include the demographic distributions for students who responded on the DAPVU.

Table 29

Distribution of Students' Demographic Variables Treated as Categorical

Variables	Total (*N*=597)		DAPVU (*n*=43)	
	n	%	*n*	%
Gender				
Male	142	23.8	5	11.6
Female	325	54.4	38	88.4
Missing	130	21.8	0	0.0
Race/Ethnicity				
American Indian or other Native American	4	0.7	0	0.0
Asian, Asian American, or Pacific Islander	9	1.5	0	0.0
Native Hawaiian	2	0.3	1	2.3
Black or African American, Non-Hispanic	54	9.0	5	11.6
White, Non-Hispanic	119	19.9	19	44.2
Hispanic, Latino, or Spanish	124	20.8	16	37.2
Other/Multi-racial	22	3.7	1	2.3
Missing	263	44.1	1	2.3
Enrollment Status during Fall 2015				
Part time enrollment	130	21.8	11	25.6
Full time enrollment	199	33.3	29	67.4
Missing	268	44.9	3	7.0
Number of Dependents				
None	232	38.9	24	55.8
1 dependent	32	5.4	5	11.6
2 dependents	33	5.5	6	14.0
3 or more dependents	28	4.7	6	14.0
Missing	272	45.6	2	4.7
English as Native (First) Language				
English is native language	291	48.7	36	83.7
English is not native language	28	4.7	5	11.6
Missing	278	46.6	2	4.7
First Generation College Student				
Participant is first generation college student	160	26.8	18	41.9
Participant is not first generation college student	159	26.6	23	53.5
Missing	278	46.6	2	4.7

Table 30

Distribution of Students' Demographic Variables Treated as Continuous

Variables	Total (N=597)		DAPVU (n=43)	
	n	%	*n*	%
Age (in years)				
Under 18 (0)	8	1.3	0	0.0
18 to 19 (1)	257	43.0	20	46.5
20 to 21 (2)	45	7.5	1	2.3
22 to 24 (3)	32	5.4	4	9.3
25 to 29 (4)	40	6.7	4	9.3
30 to 39 (5)	50	8.4	10	23.3
40 to 49 (6)	23	3.9	2	4.7
50 to 64 (7)	12	2.0	2	4.7
65 and older (8)	1	0.2	0	0.0
Missing	129	21.6	0	0.0
Job Status during Fall 2015 (hours per week)				
Not employed (0)	119	19.9	19	44.2
Less than 20 (1)	48	8.0	6	14.0
20 to 25 (2)	50	8.4	5	11.6
26 to 30 (3)	16	2.7	2	4.7
31 to 35 (4)	29	4.9	3	7.0
36 to 39 (5)	18	3.0	3	7.0
40 or more (6)	53	8.9	4	9.3
Missing	264	44.2	1	2.3
Mother's Highest Level of Education				
Didn't graduate (0)	40	6.7	6	14.0
HS or GED (1)	107	17.9	18	41.9
Some college (2)	75	12.6	6	14.0
Associate (3)	42	7.0	6	14.0
Bachelor's (4)	26	4.4	2	4.7
Master's (5)	14	2.3	0	0.0
Doctorate (6)	2	0.3	0	0.0
Unknown/other (7)	19	3.2	3	7.0
Missing	272	45.6	2	4.7

Note. Corresponding scale values for the analyses are listed in parentheses.
Note. I coded the 19 unknown/other responses for maternal education as missing for the analyses.

Table 31

Minimums, Maximums, Means, and Standard Deviations for Students' Demographic Variables Treated as Continuous

Variables	Min	Max	*M*	*SD*
Age (*n*=468)	0 (>18 years)	8 (65+ years)	2.31	1.84
Job status (*n*=333)	0 (unemployed)	6 (40+ hours/week)	2.16	2.23
Maternal Education (*n*=306)	0 (no HS degree)	6 (doctorate)	1.86	1.36

Note. Corresponding scale minimum and maximum values are provided. For "Maternal Education," the 19 participants that chose "Unknown/other" are excluded from the results presented in this table.

Appendix J: Pearson Correlation Matrices for DM-Noncognitive Factors

Tables 32, 33, and 34 include the Pearson correlations for the dm-noncognitive factor scores on the pre-, post-, and then-surveys, respectively.

Table 32

Pearson Correlation Matrix for DM-Noncognitive Factor Scores on the Pre-Survey

Variables	1	2	3	4
1 Equanimity pre-score				
2 Mindset pre-score	.22*			
3 Self-Efficacy pre-score	.18*	.46*		
4 Math Belongingness pre-score	.30*	.29*	.40*	
5 College Belongingness pre-score	.22*	.21*	.36*	.60*

Note. *$p < .01$ (2-tailed)

Table 33

Pearson Correlation Matrix for DM-Noncognitive Factor Scores on the Post-Survey

Variables	1	2	3	4
1 Equanimity post-score				
2 Mindset post-score	.23*			
3 Self-Efficacy post-score	.33*	.51*		
4 Math Belongingness post-score	.23*	.38*	.50*	
5 College Belongingness post-score	.24*	.28*	.42*	.64*

Note. *$p < .01$ (2-tailed)

Table 34

Pearson Correlation Matrix for DM-Noncognitive Factor Scores on the Then-Survey

Variables	1	2	3	4
1 Equanimity then-score				
2 Mindset then-score	.26*			
3 Self-Efficacy then-score	.36*	.58*		
4 Math Belongingness then-score	.30*	.42*	.48*	
5 College Belongingness then-score	.26*	.41*	.44*	.71*

Note. *p < .01 (2-tailed)

Appendix K: Research Question 2 Results with Transformed Attendance

Research Question 2: Do beginning-of-semester to end-of-semester differences in students' math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness correlate with semester outcomes?

> *RQ2 Research Hypothesis:* Beginning-of-semester to end-of-semester differences in students' math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness will modestly correlate with semester outcomes.

This section includes results from analyses of Research Question 2, using the transformed attendance variable instead of the original variable. The original variable was negatively skewed, but I wanted to use the fact that natural logarithm transformations can make a positively skewed distribution more normal. As such, I first reflected the variable and then took the natural logarithm: To reflect the variable, I subtracted each student's score from 101 because the maximum value for attendance is 100 and the natural logarithm function is defined only for numbers greater than 0. Then I took the natural logarithm of each student's score.

To address this research question, I used the Transformed Attendance as my dependent variable and used the change scores in dm-noncognitive factors (post-survey minus pre-survey for each of: math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness) as my independent variables. My model controls for pre-survey scores and demographic variables: gender, age, race/ethnicity, enrollment status, job status, number of dependents, native language, maternal education, and first generation college student. I initially treated section as a random effect in the model. Setting my alpha level to .017, the random effect was not significant and I did not include section as a random effect in the final model.

The results for my variables of interest (math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness) and the equanimity pre-survey score variable using the transformed attendance variable are the same as the results using the original attendance variable. The results of the two analyses differ only in results for self-efficacy pre-survey score and number of dependents; these control variables were only significant when using the transformed attendance variable. Table 35 lists results from the tests of fixed effects for the model used to address Research Question 2 with Transformed Attendance. Table 35 includes the significant and non-significant variables of interest, as well as the significant control variables. Control variables that were not significant for a model are excluded from the table, $p > .017$.

Table 35

Tests of Fixed Effects for Variables of Interest and Significant Control Variables in Research Question 2 Model with Transformed Attendance

Variables	df_1, df_2	F	p
Equanimity change	1, 220.00	7.46	.007
Mindset change	1, 220.00	0.53	>.017
Self-Efficacy change	1, 220.00	1.27	>.017
Math Belonging change	1, 220.00	0.00	>.017
College Belonging change	1, 220.00	0.16	>.017
Equanimity pre-survey	1, 220.00	6.59	.011
Self-Efficacy pre-survey	1, 220.00	5.89	.016
Dependents	1, 220.00	7.16	.008

Note. This table includes results for both significant and non-significant variables of interest for Transformed Attendance—change scores for math equanimity, math mindset, math self-efficacy, math belongingness, and college belongingness—as well as significant control variables. In the model for transformed attendance, n=243. Significance was set at $p < .017$.

There was a significant effect of equanimity change scores on students' percent attendance. Table 36 lists the coefficients and standard errors for the significant variables in the model for Transformed Attendance.

Table 36

Coefficients and Standard Errors of Significant Variables in Research Question 2 Model with Transformed Attendance

Variables	Coefficient	*SE*
Equanimity change	-0.24	0.09
Equanimity pre-survey	-0.26	0.10
Self-Efficacy pre-survey	0.28	0.12
No dependents	0.55	0.21

Note. Significance was set at $p < .017$.

Appendix L: Key Elements of Explicit Place Value Understanding

Rusch (1997) and Hannigan (1998) defined a list of twelve key elements of explicit place value understanding. Rusch (p. 97) describes the crucial elements as follows:

1. Knowledge that the single digit symbols used to represent quantity have no inherent magic or meaning – they are societal convention.
2. Knowledge that the way that people symbolically represent quantities is guided by the grouping strategies that society chooses to use for a given situation.
3. Knowledge that it is the choice to use grouping strategies that makes the system of symbolic notation so efficient and useful.
4. Knowledge that the base ten grouping strategy is a societal convention. There are a wide variety of grouping strategies that can be used – and in fact <u>are</u> used – to represent quantities symbolically.
5. Knowledge that the total value of a single digit is dependent on the symbol's location in the number as a whole.
6. Knowledge that the total quantity represented by a single digit is the product of the quantity represented by that digit alone and the quantity represented by the position of that digit.
7. Knowledge that the cardinality of the number as a whole is the sum of the quantities represented by each digit in the number.
8. Knowledge of how the algorithms for addition, subtraction, multiplication, and division emerge logically from the notational strategies used to represent quantity.
9. Knowledge of how to symbolically represent a given quantity using a variety of grouping strategies.

10. Ability to articulate the connections (similarities and differences) among positional notation structures that use systematic and non-systematic grouping schemes.
11. Ability to recognize and identify positional notation strategies (i.e., representations of quantity using a "place value" structure) using familiar, unfamiliar, systematic, and non-systematic grouping schemes.
12. Ability to modify algorithms used with familiar-systematic grouping schemes and apply those modified algorithms creatively and confidently in place value environments that use familiar-nonsystematic, unfamiliar-systematic, and unfamiliar-nonsystematic grouping schemes.

Appendix M: Developmental Assessment of Place Value Understanding

MATH BEHAVIORS STUDY: Math Assessment

Thank you for participating in the math behaviors study! There are 13 questions based on 5 problem situations. Please:

1. ***Do not work with other people**. Your responses are confidential and will help me understand the impact of your math class.*
2. ***Do not take more than 1 hour**, even if you are not finished. Just go to the last page and hit the **Submit** button.*
3. ***Include as much detail as possible** so I can understand what you are thinking.*

Gift cards will be emailed to you during the week of December 20th. If you have questions for me, you can email me at speacock@math.utexas.edu.

Thanks!
Stephanie

Problem Situation 1: Pat's Skiing Competition

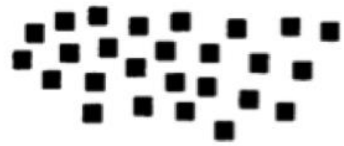
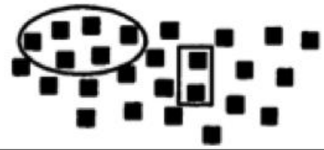
Pat is in a skiing competition. In the competition, everyone skis down the hill twice. The times are added together and the skier with the fastest combined time is the winner of the competition. Pat skied down the hill the first time in ***2 minutes and 53.67 seconds (2:53.67)****. His second time down the hill was* ***2:50.54****.*

Question 1:
What is Pat's <u>*combined*</u> *time for the competition?*

Question 2:
Describe, **in detail***, exactly what you did to solve the problem. Please be very specific because I am not able to see your work.*

**If you are not sure how to find the combined time, what ideas do you have about the problem? Are there any questions you would like answered to help you solve the problem?*

Problem Situation 2: Bobby's Squares

Mrs. Jones is working with Bobby, one of her students. Mrs. Jones draws twenty-six squares on the board and asks Bobby to write the number of squares.	
Bobby writes this:	*2 6*
Next, Mrs. Jones points to the six in the number Bobby wrote and asks him to draw a circle around the squares represented by the six.	*2 6* 
Bobby does this:	
Next, Mrs. Jones points to the two in the number Bobby wrote and asks him to draw a box around the squares represented by the two.	*2 6* 
Bobby does this:	

Question 3:
What is your evaluation of Bobby's thinking? What do you think Bobby understands?

Question 4:
Is there anything Bobby does not understand about the problem? If yes, what does Bobby not understand?

Problem Situation 3: Chocolate Factory

You recently toured a chocolate factory and found out that four is an important number for the workers. When they prepare the chocolates for shipping, they put four single chocolates in a package, four packages of chocolates in a box, four boxes of chocolates in a carton, and four cartons of chocolates in a case.

4 singles = 1 package
4 packages = 1 box
4 boxes = 1 carton
4 cartons = 1 case

The workers have developed a system so they know how many more chocolates, in the different types of containers, are needed to complete a case. Their notation is:

(# of cases)(# of cartons)(# of boxes)(# of packages)(# of singles)

Examples:

In factory notation, a 1231 is a partially filled case that has 1 carton, 2 boxes, 3 packages, and 1 single chocolate in it. This notation could mean that those containers are already in the case or the equivalent quantity of single chocolates has already been made and is just waiting to be packaged.

In factory notation, a 10000 stands for a full case that is ready to be shipped.

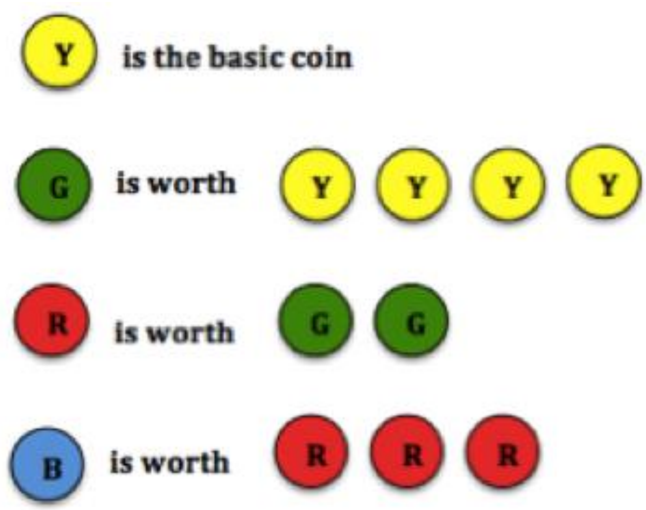
Question 5:
Inside the chocolate factory you heard someone yell, "I have a case that is partially full! It is a 3012!" How many chocolates, <u>in factory notation</u>, do they need to make the case full?

Question 6:
*Describe, **in detail**, exactly what you did to solve the problem. Please be very specific because I am not able to see your work.*

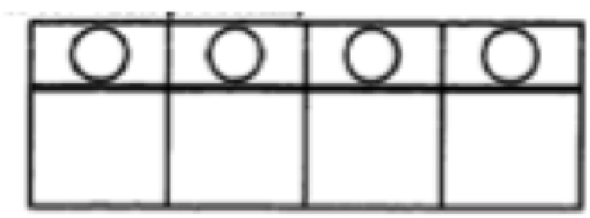
**If you are not sure how to solve the problem, what ideas do you have about the problem? Are there any questions you would like answered to help you solve the problem?*

Rugolia

The Kingdom of Rugolia has developed a money system for buying and selling merchandise. The Royal Yellow  *is the basic coin. The other coins, the Royal Green* *, the Royal Red* *, and the Royal Blue* *, are worth more. The system is the following:*



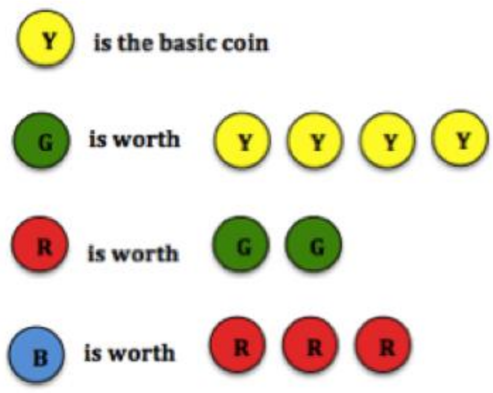
The Rugolians often use counting tables like the one below to help with calculations. Feel free to use a table if you think it would be helpful. I can't see what you are doing, so please explain what you do as much as possible.

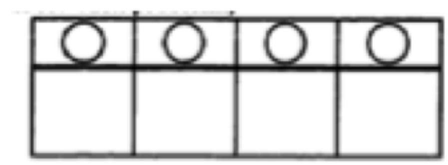


****The information about what each coin is worth will be provided with the problem.***

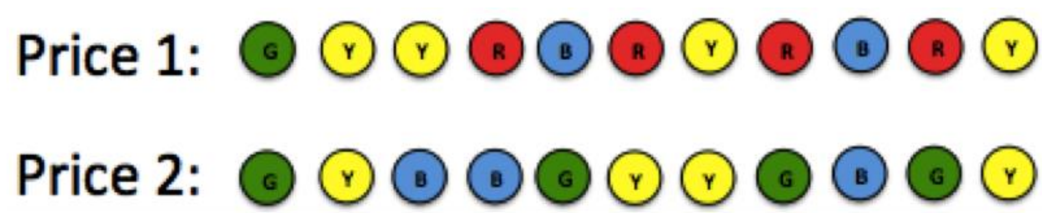
Problem Situation 4: Rugolian Rug Merchant

RUGOLIAN MONEY and COUNTING TABLE:





You are selling Rugolian rugs and you have two rugs for sale. The prices for the two rugs are:



A buyer wants to buy both rugs and asks for your price using the fewest number of coins possible.

Question 7:
What is the price for the pair using the fewest number of coins possible?

Question 8:
Describe, ***in detail****, exactly what you did to solve the problem. Please be very specific because I am not able to see your work.*

**If you are not sure how to solve the problem, what ideas do you have about the problem? Are there any questions you would like answered to help you solve the problem?*

Problem Situation 5: Maria's Subtraction

Maria is a great student, but she is having some trouble with subtraction. Below is an example of Maria's subtraction work:

$$\begin{array}{rrrr} & & 8 & \\ & 6 & \cancel{9} & {}^{1}3 \\ - & 2 & 4 & 8 \\ \hline & 4 & 4 & 5 \end{array} \qquad \begin{array}{rrrr} & 5 & & \\ & \cancel{7} & {}^{1}2 & {}^{1}6 \\ - & 3 & 4 & 9 \\ \hline & 2 & 8 & 7 \end{array} \qquad \begin{array}{rrrr} & 2 & & \\ & \cancel{4} & {}^{1}3 & {}^{1}4 \\ - & 2 & 7 & 6 \\ \hline & & 6 & 8 \end{array} \qquad \begin{array}{rrrr} & 2 & & \\ & \cancel{3} & {}^{1}2 & 5 \\ - & 1 & 5 & 1 \\ \hline & 1 & 7 & 4 \end{array}$$

Here are three problems Maria needs to answer:

$$\begin{array}{rrrr} & 4 & 5 & 1 \\ - & 2 & 3 & 9 \\ \hline \end{array} \qquad \begin{array}{rrrr} & 7 & 3 & 0 \\ - & 2 & 8 & 4 \\ \hline \end{array} \qquad \begin{array}{rrrr} & 8 & 5 & 3 \\ - & 4 & 9 & 7 \\ \hline \end{array}$$

Maria's errors indicate that there is something she does not understand.

Question 9:
What answer would Maria give for this problem?

$$\begin{array}{rrrr} & 4 & 5 & 1 \\ - & 2 & 3 & 9 \\ \hline \end{array}$$

Question 10:
What answer would Maria give for this problem?

$$\begin{array}{rrrr} & 7 & 3 & 0 \\ - & 2 & 8 & 4 \\ \hline \end{array}$$

Question 11:
What answer would Maria give for this problem?

$$\begin{array}{rrrr} & 8 & 5 & 3 \\ - & 4 & 9 & 7 \\ \hline \end{array}$$

Question 12:
Describe, ***in detail****, exactly what you did to come up with the answers you think Maria would give for the problems. Please be very specific because I am not able to see your work.*

Question 13:
Describe your assessment of what concept or concepts Maria does not understand. Try to explain what she does not understand, not just what procedure she is using.

This is the end of the assessment. Hit the Next Button if you would like to submit your responses. You cannot open the assessment once you have submitted your responses. Thank you for your time! I hope you found this interesting!

Appendix N: DAPVU Rubrics

The Developmental Assessment of Place Value Understanding rubrics are slightly modified versions of Rusch (1997) and Hannigan's (1998) APVU rubrics.

Problem Situation 1: Pat's Skiing Competition

Objective: To assess the ability to recognize a simple variation of a place value situation and to adapt the traditional computation algorithms accordingly.
Evidence: Sophistication of discernible strategies indicated in the written calculations.
Mathematical Task: Addition in a nonsystematic base in a familiar context.
Knowledge Dimension Rubrics:

Accurate Computation

Level 5: Correct computation.
Level 4: Incorrect computation caused by a computational error.
Level 3: Incorrect computation with evidence of a minor conceptual error.
Level 2: Incorrect computation with evidence of significant conceptual errors, incomplete computation, or total confusion.
Level 1: No attempt at task.

Analysis of Computation Method

Level 5: Complete, sophisticated, and insightful adaptation of the traditional algorithm within the mixed-grouping place value structure. For example, the computation process uses only an adaptation of the traditional algorithm; i.e., symbolic regrouping is used accurately across units (minutes, seconds, hundredths of seconds) as well as within units.
Level 4: Partial adaptation of the traditional algorithm within the mixed grouping place value structure. For example, symbolic regrouping may be used from hundredths of seconds to seconds, but not used across larger units; instead, an appropriate alternative regrouping strategy is used.
Level 3: No evidence of the adaptation of the traditional algorithm; however, alternative regrouping strategies are consistently applied to the mixed-grouping place value structure. OR A partial adaptation was utilized but the alternative regrouping strategy was left incomplete.
Level 2: No evidence of the adaptation of the traditional algorithm to the mixed-grouping place value structure. Alternative regrouping strategies may have been attempted but are disorganized and/or inaccurately applied.
Level 1: No evidence that the mixed-grouping place value structure is recognized or utilized. Base-ten strategies may have been consistently applied in inappropriate situations. OR Computation (task) is incomplete with insufficient evidence to determine a strategy.

<u>Problem Situation 2: Bobby's Squares</u>

Objective: To measure explicit understanding of notational structure.
Evidence: The depth of analysis indicated in the verbal description of the child's thinking/understanding and sophistication of the place value language used.
Mathematical Task: Notational structure of a systematic base in a familiar context.
Knowledge Dimension Rubrics:

Depth of Analysis

Level 5: Develops an accurate and elaborate analysis of the place value concepts not understood by the child.

Level 4: Develops an accurate analysis of the place value concepts not understood by the child.

Level 3: Mentions what is not understood by the child, but leaves it undeveloped. OR Evidence suggests that the analysis provided, which is an accurate analysis, is not offered as a reasonable analysis but as one possible alternative analysis (the analysis is not recognized as the accurate one.)

Level 2: Provides an accurate description of some or all behaviors, but no analysis of understanding.

Level 1: Provides an analysis that is irrelevant, incorrect, or uninformative.

Use of Descriptive Language

Level 5: Uses accurate and highly specific place value language. For example, specifically articulates that there is an association between the digit 2 and two groups of ten.

Level 4: Uses accurate and specific place value language. For example, specifically articulates that there is an association between the digit 2 and twenty.

Level 3: Uses accurate but non-specific place value language to describe a place value concept. For example, the association between the digit 2 and twenty or two groups of ten is implied rather than clearly articulated.

Level 2: May use accurate but non-specific place value language; however, evidence suggests that the language is used to indicate an observed behaviors rather than to describe a place value concept of "groups of tens."

Level 1: No or inaccurate use of place value language. OR Analysis does not use place value language to describe behaviors.

Problem Situation 3: Chocolate Factory

Objective: To assess the ability to recognize a simple variation of a place value situation and to adapt the traditional computation algorithms accordingly.
Evidence: Sophistication of discernible strategies indicated in the written calculations.
Mathematical Task: Subtraction in a systematic base in an unfamiliar context.
Knowledge Dimension Rubrics:

Accurate Computation

- Level 5: Correct computation shown in factory notation.
- Level 4: Incorrect computation caused by a computational error.
- Level 3: Incorrect computation with evidence of a minor conceptual error.
- Level 2: Incorrect computation with evidence of significant conceptual errors, incomplete computation, or total confusion.
- Level 1: No attempt at task. OR The attempt is irrelevant or inappropriate. OR Evidence suggests that there was a misinterpretation of the task.

Analysis of Computation Method

- Level 5: Complete, sophisticated, and insightful adaptation of the traditional base-ten subtraction algorithm to the base-n place value structure. There is no indication of a need to use illustrations, diagrams, charts, etc. to clarify and/or support the computation.
- Level 4: Reasonable adaptation of a base-ten subtraction algorithm within the base-n place value structure; for example, a well utilized adding-up scheme. There may be illustrations, diagrams, charts, etc. used to clarify and/or support the computation.
- Level 3: No evidence of a reasonable adaptation of a base-ten subtraction algorithm. However, alternative symbolic and/or pictorial regrouping strategies were consistently applied to the base-n place value structure; for example, accurate conversion to base-ten followed by subtraction and conversion back to base-n. OR Poorly executed but reasonable adaptation of a base-ten subtraction algorithm.
- Level 2: No evidence of any adaptation of a base-ten subtraction algorithm. Alternative symbolic and/or pictorial regrouping strategies may have been attempted but are disorganized and/or inaccurately applied to the base-n place value structure. OR Incomplete alternative symbolic and/or pictorial regrouping strategy.
- Level 1: No evidence that the base-n place value structure was recognized or utilized. Base-ten strategies may have been consistently applied in inappropriate situations. OR Computation (task) is incomplete with insufficient evidence to determine a strategy.

Problem Situation 4: Rugolian Rug Merchant

Objective: To assess the ability to recognize a simple variation of a place value situation and to adapt the traditional computation algorithms accordingly.
Evidence: Sophistication of discernible strategies indicated in the written calculations.
Mathematical Task: Addition in a nonsystematic base in an unfamiliar context.
Knowledge Dimension Rubrics:

Accurate Computation

Level 5: Correct computation.
Level 4: Incorrect computation caused by a computational error.
Level 3: Incorrect computation with evidence of a minor conceptual error.
Level 2: Incorrect computation with evidence of significant conceptual errors, incomplete computation, or total confusion.
Level 1: No attempt at task (no work shown).

Analysis of Computation Method

Level 5: Complete, sophisticated, and insightful symbolic adaptation of the traditional base-ten algorithms to the mixed-grouping place value structure. For example, symbolic (digits rather than pictures) regrouping is part of the algorithm. There is no indication of a need to use illustrations.
Level 4: Partial symbolic adaptation of the traditional base-ten algorithms to the mixed grouping place value structure. There may be illustrations used to clarify and/or support the computation.
Level 3: No evidence of any symbolic adaptation of the traditional base-ten algorithms; however, alternative symbolic and/or pictorial strategies were consistently applied to the mixed-grouping place value structure. For example, regrouping the quantity as all Ys, accurate computation (or with minor error), and regrouping as fewest number of coins is an appropriate strategy.
Level 2: No evidence of any symbolic adaptation of the traditional base-ten algorithms. Alternative symbolic and/or pictorial strategies may have been attempted but are disorganized and/or inaccurately applied to the mixed-grouping place value structure. For example, regrouping the quantity as all Ys that has major computational errors or is not regrouped as the fewest number of coins is a poor attempt at a strategy. OR The computation (task) is incomplete, through confusion or omission, but there is some clear evidence that alternative strategies were consistently applied to some elements of the task.
Level 1: No evidence that the mixed-grouping place value structure was recognized or utilized. Calculation strategies used are inappropriate situations. Computation (task) is incomplete with no evidence of

consistent application or alternative strategies; or computation (task) is not attempted.

Analysis of Symbolic Representation

The symbolic representation demonstrated should be evaluated on type of representation (symbolic, algebraic, or pictorial). The counting table may be used to represent the quantities but is not required to be used.

The given quantities may be regrouped prior to computation but that demonstration is not required. Although expression of each quantity appropriately grouped (part of the symbolic representation process) prior to computation indicates a more sophisticated understanding, it is not necessary—perhaps even inefficient—to answer the questions asked. Since it is impossible to discern, from written computations, the student's reason for NOT initially regrouping the quantity, this sophistication of understanding is not being assessed.

Level 5: Constructs a logical and consistent symbolic representation (i.e., digits only) using place value columns which are organized in either an increasing or decreasing order (i.e., Y, G, R, B or B, R, G, Y).

Level 4: Constructs a logical and consistent algebraic representation (i.e., digits and letters) using place value columns which are organized in either an increasing or decreasing order (i.e., Y, G, R, B or B, R, G, Y).

Level 3: Constructs a logical and consistent pictorial representation (i.e., circles, tally marks, or letters without digits) using place value columns which are organized in either an increasing or decreasing order (i.e., Y, G, R, B or B, R, G, Y). OR Constructs a logical and consistent symbolic or algebraic representation in which place value columns are utilized, but not in an increasing or decreasing order.

Level 2: Constructs a reasonable representation which may have algebraic or symbolic elements, but does not utilize place value columns. For example, converts to all yellow coins and uses digits to compute in base-ten, and then converts back to mixed coins.

Level 1: Attempts to construct a representation but the result of the attempt is inaccurate or incomplete.

Problem Situation 5: Maria's Error Pattern

Objective: To measure explicit understanding of the connections among notational structure, regrouping and algorithms.
Evidence: Recognition and replication of the child's thinking, the depth of analysis indicated in the verbal description of the child's thinking/understanding and sophistication of the place value language used.
Mathematical Task: Subtraction in a systematic base in a familiar context (with a standard algorithm).
Knowledge Dimension Rubrics:

Error Reproduction

Level 5: Accurate reproduction of error.
Level 4: Evidence that the error is understood but a computational error exists.
Level 3: Inaccurate reproduction of error with evidence that there is partial recognition of the error pattern.
Level 2: Inaccurate reproduction of error with no evidence that there is recognition of the error pattern. OR An incomplete attempt.
Level 1: No attempt at task.

Depth of Analysis

Level 5: Develops an accurate and elaborate analysis of the place value concepts not understood by the child.
Level 4: Develops an accurate analysis of the place value concepts not understood by the child.
Level 3: Mentions what is not understood by the child, but leaves it undeveloped.
Level 2: Provides an accurate description of some or all behaviors, but no analysis of understanding.
Level 1: Provides an analysis that is irrelevant, incorrect, or uninformative.

Use of Descriptive Language

Level 5: Uses accurate and highly specific place value language. For example, uses the formal term "regrouping" in place of the informal terms "carrying" or "borrowing" and "groups of ten" in place of "ten" or "one".
Level 4: Uses accurate and specific place value language. For example, uses the informal terms "carrying" and "borrowing" but evidence suggests that those terms are being used as synonyms for the formal term "regrouping" and/or uses the word "ten" but evidence suggests that the term is being used as a synonym for the more specific phrase "one group of ten."
Level 3: Uses accurate but non-specific place value language to describe a place value concept. For example, uses the informal terms "carrying" and "borrowing" as synonyms for the formal term "regrouping" and/or the

word "one" is being used as a synonym for the more specific word "ten" or phrase "group of ten."

Level 2: May use accurate but non-specific place value language; however, evidence suggests that the language is used to indicate an observed behavior rather than to describe a place value concept.

Level 1: No or inaccurate use of place value language. OR Analysis does not use place value language to describe behaviors.

Appendix O: Evidence Markers for the Online DAPVU

Problem Situation 1: Pat's Skiing Competition

Code	***PS1: Accuracy***
(0)	Incorrect
(1)	Correct
(99)	No computation

Code	***PS1: Representation***
(0)	Digits and units (e.g., 5 minutes, 44 seconds, 21 milliseconds; 5 minutes, 44.21 seconds)
(1)	Fully symbolic nonstandard representation (e.g., 5:44:21)
(2)	Fully symbolic standard representation (e.g., 5:44.21)
(99)	No computation

Code	***PS1: Depth of Understanding***
(0)	No evidence that the mixed-grouping place value structure is recognized or utilized. Has significant conceptual errors (e.g., "add the minutes then the hours divided by 24" (i.e., 53.67+50.54+2+2)/24)).
(1)	No evidence that the mixed-grouping place value structure is recognized or utilized. Base-ten strategies may have been consistently applied in inappropriate situations.
(2)	No evidence that the mixed-grouping place value structure is recognized or utilized. Base-sixty strategies may have been consistently applied.
(3)	Evidence that the mixed-grouping place value structure is recognized and utilized. Regrouping strategies are applied to the mixed-grouping place value structure but are disorganized and/or inaccurately applied. Has conceptual error.
(4)	Evidence that the mixed-grouping place value structure is recognized and utilized. Regrouping strategies are applied to the mixed-grouping place value structure but are disorganized and/or inaccurately applied. Has minor computational error.
(5)	Evidence that the mixed-grouping place value structure is recognized and utilized. Regrouping strategies are consistently applied to the mixed-grouping place value structure (e.g., recognition that 100 milliseconds = 1 second and 60 seconds = 1 minute). No computational errors.
(97)	Unknown - Insufficient evidence to decipher computation method
(98)	Unknown - Provides estimation
(99)	Unknown - No response

Problem Situation 2: Bobby's Squares

Code	*PS2: Descriptive Language*
(0)	No or inaccurate use of place value language. OR Analysis does not use place value language to describe behaviors.
(1)	May use accurate but non-specific place value language; however, evidence suggests that the language is used to indicate an observed behaviors rather than to describe a place value concept of "groups of tens."
(2)	Uses accurate but non-specific place value language to describe a place value concept. For example, the association between the digit 2 and twenty or two groups of ten is implied rather than clearly articulated.
(3)	Uses accurate and specific place value language. For example, specifically articulates that there is an association between the digit 2 and twenty.
(4)	Uses accurate and highly specific place value language. For example, specifically articulates that there is an association between the digit 2 and two groups of ten.

Code	*PS2: Depth (of Analysis)*
(0)	Provides an analysis that is irrelevant, incorrect, or uninformative.
(1)	Provides an accurate description of some or all behaviors, but no analysis of understanding.
(2)	Mentions what is not understood by the child, but leaves it undeveloped. OR Evidence suggests that the analysis provided, which is an accurate analysis, is not offered as a reasonable analysis but as one possible alternative analysis (the analysis is not recognized as the accurate one.)
(3)	Develops an accurate analysis of the place value concepts not understood by the child.
(4)	Develops an accurate and elaborate analysis of the place value concepts not understood by the child.
(99)	Unknown - No response

Problem Situation 3: Chocolate Factory

Code	*PS3: Accuracy*
(0)	Incorrect
(1)	Correct
(99)	No computation

Code	*PS3: Representation*
(0)	Digits and units (e.g., 3 boxes, 2 packages, 2 singles)
(2)	Fully symbolic representation with digits only (e.g., 322)
(99)	No computation

Code	*PS3: Depth (of Understanding)*
(0)	No evidence that the base-four place value structure is recognized or utilized. Base-ten strategies may have been consistently applied in inappropriate situations. Has significant conceptual errors.
(1)	Evidence that the base-four place value structure is recognized and utilized. Regrouping strategies are applied to the base-four place value structure but are disorganized and/or inaccurately applied. Has significant conceptual error (e.g., a full case = 4444).
(2)	Evidence that the base-four place value structure is recognized and utilized. Regrouping strategies are applied to the base-four place value structure but are disorganized and/or inaccurately applied. Has minor conceptual error.
(3)	Evidence that the base-four place value structure is recognized and utilized. Regrouping strategies are applied to the base-four place value structure but are disorganized and/or inaccurately applied. Has minor computational error.
(4)	Evidence that the base-four place value structure is recognized and utilized. Regrouping strategies are consistently applied to the base-four place value structure. No computational errors.
(97)	Unknown - Insufficient evidence to decipher computation method
(99)	Unknown - No response
(100)	Unknown - Restates initial quantity with units or restates example quantity
(101)	Unknown - Explicitly expresses confusion, but does not provide calculations

Problem Situation 4: Rugolian Rug Merchant

Code	*PS4: Accuracy*
(0)	Incorrect
(1)	Correct
(99)	No computation

Code	*PS4: Representation*
(0)	Digits and units (e.g., 7 blue, 1 red, 1 green)
(1)	Letters only (e.g., BBBBBBBRG)
(2)	Fully symbolic representation (e.g., 7:1:1:0)
(3)	Other (nonsensical) representation (e.g., $200)
(99)	No computation

Code	*PS4: Depth (of Understanding)*
(0)	No evidence that the mixed-grouping place value structure is recognized or utilized. Calculation strategies are used in inappropriate situations.
(1)	Evidence that the mixed-grouping place value structure is recognized and utilized. Regrouping strategies are applied to the mixed-grouping place value structure but are disorganized and/or inaccurately applied. Has conceptual or computational errors.
(2)	Evidence that the mixed-grouping place value structure is recognized and utilized. Regrouping strategies are correctly applied to the mixed-grouping place value structure with no computational errors, but regrouping is left incomplete (e.g., does not convert 3G to 1R and 1G).
(3)	Evidence that the mixed-grouping place value structure is recognized and utilized. Regrouping strategies are correctly applied to the mixed-grouping place value structure with no computational errors, but uses unnecessary regrouping strategies (e.g., initially converts quantity to all yellow coins).
(4)	Evidence that the mixed-grouping place value structure is recognized and utilized. Regrouping strategies are correctly applied to the mixed-grouping place value structure with no computational errors. Uses efficient regrouping strategies and represents quantity using the fewest number of coins possible.
(97)	Unknown - Insufficient evidence to decipher computation method
(99)	Unknown - No response
(101)	Unknown - Explicitly expresses confusion, but does not provide calculations or provides seemingly random computation
(102)	Unknown - Likely thinks problem is asking for which price is higher or lower

Problem Situation 5: Maria's Error Pattern

Code	*PS5: Accuracy*
(0)	Computation without reproduction of error pattern
(1)	Partial reproduction of error pattern
(2)	Accurate reproduction of error pattern
(99)	No computation

Code	*PS5: Descriptive Language*
(0)	No or inaccurate use of place value language. OR Analysis does not use place value language to describe behaviors.
(1)	May use accurate but non-specific place value language; however, evidence suggests that the language is used to indicate an observed behavior rather than to describe a place value concept.
(2)	Uses accurate but non-specific place value language to describe a place value concept. For example, uses the informal term "borrowing" as a synonym for the formal term "regrouping" and/or the word "one" is being used as a synonym for the more specific word "ten" or phrase "group of ten."
(3)	Uses accurate and specific place value language. For example, uses the informal term "borrowing" but evidence suggests that term is being used as a synonym for the formal term "regrouping" and/or uses the word "ten" but evidence suggests that the term is being used as a synonym for the more specific phrase "one group of ten."
(4)	Uses accurate and highly specific place value language. For example, uses the formal term "regrouping" in place of the informal term "borrowing" and "groups of ten" in place of "ten" or "one".
(99)	Unknown – Provides computation with no analysis of understanding

Code	*PS5: Depth (of Analysis)*
(0)	Provides an analysis that is irrelevant, incorrect, or uninformative.
(1)	Provides an accurate description of some or all behaviors, but no analysis of understanding.
(2)	Mentions what is not understood by the child, but leaves it undeveloped. OR Evidence suggests that the analysis provided, which is an accurate analysis, is not offered as a reasonable analysis but as one possible alternative analysis (the analysis is not recognized as the accurate one.)
(3)	Develops an accurate analysis of the place value concepts not understood by the child.
(4)	Develops an accurate and elaborate analysis of the place value concepts not understood by the child.
(99)	Unknown – Provides computation with no analysis of understanding

Appendix P: Template for the Semi-Structured Interviews

The template for the semi-structured interviews is arranged by topic. The template served only as a guide for the interviews; I did not ask every participant every question and I did not always phrase the questions in the way they are listed. I attempted to ask questions in a conversational manner and use the least judgmental wording, avoiding using too many value labels. I loosely followed the template from beginning to end, but sometimes skipped around to follow the interviewee's apparent progression.

Frameworks

- Were you in a student success course? What was it called? Frameworks? When did you take the course? Was it related to the Foundations course?

Background

- Was this your first semester in college? First math class in college? How long since your last math class?

Math Interest

- What was your perception of math before this class? Did you like math before this class? Now? What changed?

Foundations

- Can you tell me a little about your math class this semester? Just give me an overall impression what it was like. Can you describe a typical day? Did this semester feel any different from previous math experiences? How so?
- Do you feel like you learned anything valuable? Like what (math, persistence, attitudes)?
- What is the most important thing you learned? Do you think this class will make you more successful? In what ways?

Math anxiety (An adverse emotional reaction to math or the prospect of doing math)

- When you were asked to solve problems, did you have any emotional reactions? Nerves? Excitement? Etc.
- Is there anything in math that makes you particularly anxious? Being in class? Solving problems? Taking tests?
- What about when you took math tests? Was this the same or different from when you took tests in other classes?
- Did any of this change over the course of the semester? When did it change? How was it before this class? During? Now? In what ways was it different?

Math productive persistence (Navigating and persevering through mathematical situations by utilizing effective strategies and adapting ineffective strategies)

- What do you do when you can't solve a problem? How long do you work on it? Do you reference other sources (books, peers)?
- Did you ever think about giving up on problems or dropping the course?
- Did you try harder? Did you hand in more work?
- Did any of this change over the course of the semester? When did it change? How was it before this class? During? Now? In what ways was it different?

Math belongingness (Belief that one is an accepted member of the math community whose presence and contributions are valued)

- How comfortable did you feel working with others in your class?
- Did you feel like a member of the group? Did others appreciate your contributions? Did you feel like your ideas were valued?
- How about at the college, in general?
- Did any of this change over the course of the semester? When did it change? How was it before this class? During? Now? In what ways was it different?

Math growth mindset (Belief that one's math intelligence is malleable through experiences & effort)

- Do you think there is such a thing as a math person? Do you consider yourself a math person? When did this change?
- When you are solving problems, do you think more of what others will think? External goals? Are you just trying to figure things out for personal satisfaction?
- Did any of this change over the course of the semester? When did it change? How was it before this class? During? Now? In what ways was it different?

Math transfer (Degree to which one can apply one's math knowledge & is prepared to learn in novel math situations)

- Do you think this class has prepared you to take other math classes? Use math in the real world? How so?

Math self-efficacy (Belief in one's capabilities to complete particular math tasks or meet math course objectives)

- If I gave you a math problem right now, how successful do you think you would be solving it? What if I gave you extra time and resources? (Can show the tasks)
- What types of problems do you feel most comfortable solving?
- Imagine you failed on a particular type of problem. What do you do after that? The next time you encounter that type of problem, how do you feel?
- Did any of this change over the course of the semester? When did it change? How was it before this class? During? Now? In what ways was it different?

Impact of failure (Attributions)

- In general, when you don't do as well as you would have liked solving a math problem what do you think the reason is? (Tired? Stressed? Lack of effort? Lack of innate ability?)

Think-Aloud (DAPVU and modified APVU problem situations)

- (If not sure how to solve the problem) What ideas do you have about the problem? Are there any questions you would like answered to help you solve the problem?
- Why did you choose to solve it that way? Are there other ways you could solve this problem? How would you approach a problem like this in your math class?
- What about the problem is difficult?
- Do you think you could solve it in another situation? At home alone? With resources? With enough time?
- If you were given unlimited time to solve it, do you think you would keep working on it?

Survey

- When you took the surveys in class, had you already covered some of the topics that were addressed in the survey?
- Was the survey easy or difficult to understand?

Grades

- How did you do in your class this semester? Homework? Quizzes? Exams? Final? Course grade?

Problem Situations included in the Think-Aloud

- DAPVU online problem situations
- DAPVU think-aloud problem situations:
    - *Mark's Addition (based on APVU's familiar-systematic Juan's Addition Error Pattern problem)*
    - *Chocolate Factory (based on APVU's unfamiliar-systematic Caramel Factory conversion problem)*
    - *Space Shuttle (based on APVU's familiar-nonsystematic Space Shuttle subtraction problem)*
    - *Rugolia (based on APVU's unfamiliar-nonsystematic Zorandria multiplication problem)*

Think-Aloud: Mark's Addition

Mark is a good student, but he is having trouble with addition. Below is a sample of Mark's addition work:

$$\begin{array}{r} {\scriptstyle 1} \\ 8\ 6 \\ +\ 3\ 1 \\ \hline 1\ 8 \end{array} \qquad \begin{array}{r} 5\ 1\ 7 \\ +\ 3\ 6\ 2 \\ \hline 8\ 7\ 9 \end{array} \qquad \begin{array}{r} {\scriptstyle 6} \\ 2\ 9\ 4 \\ +\ 3\ 7\ 6 \\ \hline 5\ 1\ 1\ 6 \end{array} \qquad \begin{array}{r} {\scriptstyle 0\ 1} \\ 6\ 3\ 2 \\ +\ 4\ 8\ 1 \\ \hline 1\ 1\ 4 \end{array}$$

Task: *Here are three problems Mark needs to answer. <u>Try to solve the problems how you think Mark would solve them</u>. Think aloud as you work through the problems, describing in detail exactly what you are doing. Please be very specific.*

$$\begin{array}{r} 7\ 2\ 1 \\ +\ 4\ 6\ 2 \\ \hline \end{array} \qquad \begin{array}{r} 3\ 4\ 1 \\ +\ 5\ 2\ 6 \\ \hline \end{array} \qquad \begin{array}{r} 8\ 2\ 9 \\ +\ 6\ 7\ 4 \\ \hline \end{array}$$

Task: *Mark's errors show that there is a concept he does not understand. Describe your assessment of what concept or concepts Mark does not understand. Try to explain what he does not understand, not just what procedure he is using.*

Think-Aloud: Chocolate Factory

You recently toured a chocolate factory and found out that four is an important number for the workers. When they prepare the chocolates for shipping, they put four single chocolates in a package, four packages of chocolates in a box, four boxes of chocolates in a carton, and four cartons of chocolates in a case.

4 singles = 1 package
4 packages = 1 box
4 boxes = 1 carton
4 cartons = 1 case

The workers have developed a system so they know how many more chocolates, in the different types of containers, are needed to complete a case. Their notation is:

(# of cases)(# of cartons)(# of boxes)(# of packages)(# of singles)

Examples:

In factory notation, a 1231 is a partially filled case that has 1 carton, 2 boxes, 3 packages, and 1 single chocolate in it. This notation could mean that those containers are already in the case or the equivalent quantity of single chocolates has already been made and is just waiting to be packaged.

In factory notation, a 10000 stands for a full case that is ready to be shipped.

Task: *Before you left the factory, the workers gave you thirty packages of chocolates. How would you represent thirty packages* *in factory notation*? *Think aloud as you work through the problem, describing in detail exactly what you are doing. Please be very specific.*

Think-Aloud: Space Shuttle

When the people at NASA launch a space shuttle, they want to make sure all of the things that need to happen are perfectly coordinated. They keep track of the timing of events down to fractions of seconds. The way they keep track of the time of day looks like this:

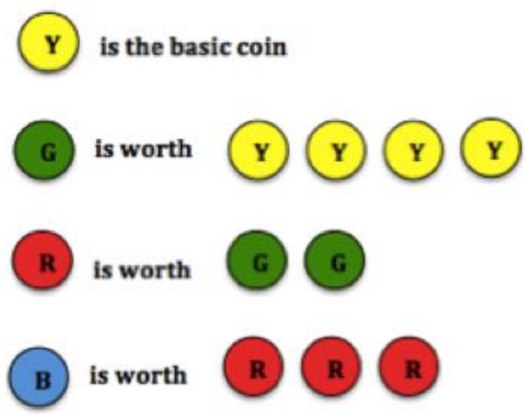
Hours : Minutes : Seconds : Hundredths of a Second

*The shuttle first fired its thrusters at **8:44:35:16** on Friday morning. The next time it fired its thrusters was at **11:26:33:07** on the same Friday morning.*

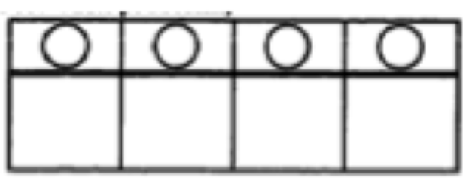
Task: *How much time elapsed in between one time it fired its thrusters and the next time it fired its thrusters? Think aloud as you work through the problem, describing in detail exactly what you are doing. Please be very specific.*

Think-Aloud: Rugolia

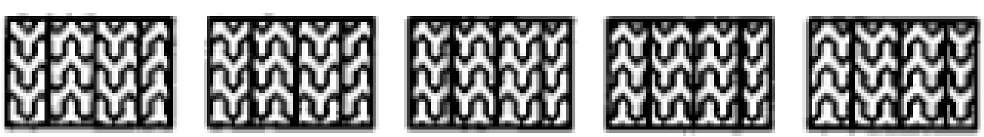
The Kingdom of Rugolia has developed a money system for buying and selling merchandise. The Royal Yellow  *is the basic coin. The other coins, the Royal Green*  *, the Royal Red*  *, and the Royal Blue*  *, are worth more. The system is the following:*



The Rugolians often use counting tables like the one below to help with calculations. Feel free to use a table if you think it would be helpful. I can't see what you are doing, so please explain what you do as much as possible.



A Rugolian rug merchant has this many rugs for sale in one area of her shop:



Each rug costs this much:

A buyer wishes to purchase all of them.

Task: *Using the fewest number of coins possible, what is the price for all these rugs? Feel free to use one or more counting tables if you think it will be helpful. Think aloud as you work through the problem, describing in detail exactly what you are doing. Please be very specific.*

References

Abts, M. (2012). *Effectiveness of online community college success courses*. Arizona State University. Retrieved from ProQuest Dissertations & Theses Full Text.

Achieving the Dream, American Association of Community Colleges, Charles A. Dana Center, Complete College America, Education Commission of the States, & Jobs for the Future. (2015a). *Core principles for transforming remediation within a comprehensive student success strategy--A joint statement*. Retrieved from http://www.core-principles.org

Achieving the Dream, American Association of Community Colleges, Charles A. Dana Center, Complete College America, Education Commission of the States, & Jobs for the Future. (2015b). *National organizations and states endorse design principles to support student success and scale effective higher education practices*. Retrieved from http://www.core-principles.org

Akin, A., & Kurbanoglu, I. N. (2011). The relationships between math anxiety, math attitudes, and self-efficacy: A structural equation model. *Studia Psychologica*, *53*(3), 263–273.

Allen, J. M., & Nimon, K. (2007). Retrospective pretest: A practical technique for professional development evaluation. *Journal of Industrial Teacher Education*, *44*(3), 27–42.

American Association of Community Colleges. (2012). *Reclaiming the American dream: A report from the 21st-Century Commission on the Future of Community*

Colleges. Washington, D.C. Retrieved from http://www.aacc.nche.edu/21stCenturyReport

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science*, *11*(5), 181–185. https://doi.org/10.1111/1467-8721.00196

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, *130*(2), 224. https://doi.org/10.1037//0096-3445.130.2.224

Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review*, *14*(2), 243–248. https://doi.org/10.3758/BF03194059

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1978). The self system in reciprocal determinism. *American Psychologist*, *33*(4), 344–358. https://doi.org/10.1037/0003-066X.33.4.344

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*(2), 122–147. https://doi.org/10.1037/0003-066X.37.2.122

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W. H. Freeman & Co.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, *52*(1), 1–26. https://doi.org/10.1146/annurev.psych.52.1.1

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology*, *41*(3), 586–598. https://doi.org/10.1037/0022-3514.41.3.586

Beach, K. (1999). Consequential transitions: A sociocultural expedition beyond transfer in education. *Review of Research in Education*, *24*, 101–139. https://doi.org/10.2307/1167268

Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: Working memory and "choking under pressure" in math. *Psychological Science*, *16*(2), 101–105. https://doi.org/10.2307/40064185

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences*, *107*(5), 1860–1863. https://doi.org/10.1073/pnas.0910967107

Berg, B. (2000). *Qualitative research methods for the social sciences* (4th ed.). Needham Heights, MA: Allyn & Bacon. Retrieved from http://www.scribd.com/doc/66314387/Qualitative-Research-Methods-for-the-Social-Sciences-4th-Edition

Betz, N. E. (1978). Prevalence, distribution, and correlates of math anxiety in college students. *Journal of Counseling Psychology*, *25*(5), 441–448. https://doi.org/10.1037/0022-0167.25.5.441

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*(1), 246–263. https://doi.org/10.1111/j.1467-8624.2007.00995.x

Bogdan, R. C., & Biklen, S. K. (2003). *Qualitative research for education: An introduction to theories and methods* (4th ed.). New York, NY: Pearson Education group.

Borghans, L., Duckworth, A. L., Heckman, J. J., & Weel, B. ter. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, *43*(4), 972–1059. https://doi.org/10.1353/jhr.2008.0017

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, 61–100.

Bray, J. H., Maxwell, S. E., & Howard, G. S. (1984). Methods of analysis with response-shift bias. *Educational and Psychological Measurement*, *44*(4), 781–804. https://doi.org/10.1177/0013164484444002

Britten, N. (1995). Qualitative interviews in medical research. *BMJ*, *311*, 251–253. https://doi.org/10.1136/bmj.311.6999.251

Broudy, H. S. (1977). Types of knowledge and purposes of education. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J. : New York: Erlbaum. Retrieved from http://catalog.hathitrust.org/Record/004425708

Brown, I., & Inouye, D. (1978). Learned helplessness through modeling: The role of perceived similarity in competence. *Journal of Personality and Social Psychology*, *36*, 900–908. https://doi.org/doi.apa.org/journals/psp/36/8/900

Brummelman, E., & Walton, G. M. (2015). “If you want to understand something, try to change it”: Social-psychological interventions to cultivate resilience. *Behavioral and Brain Sciences*, *38*, e96. https://doi.org/10.1017/S0140525X14001472

Bryk, A., Yeager, D. S., Hausman, H., Muhich, J., Dolle, J., Grunow, A., … Gomez, L. (2013). *Improvement research carried out through networked improvement communities: Accelerating learning about practices that support more productive student mindsets* (White paper prepared for the White House meeting on “Excellence in education: The importance of academic mindsets”). Retrieved from http://cdn.carnegiefoundation.org/wp-content/uploads/2014/09/improvement_research_NICs_bryk-yeager.pdf

Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin Company.

Campbell, R. (2006, March 27). Jean Piaget’s genetic epistemology: Appreciation and critique. Retrieved from http://hubcap.clemson.edu/~campber/piaget.html

Carraher, D., & Schliemann, A. (2002). The transfer dilemma. *The Journal of the Learning Sciences*, *11*(1), 1–24.

Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (2000). Mathematics in the streets and in schools. In P. Smith & A. D. Pellegrini (Eds.), *Psychology of education: Major themes* (Vol. 3, pp. 239–250). New York, NY: RoutledgeFalmer.

Carraher, T. N., & Schliemann, A. D. (1985). Computation routines prescribed by schools: Help or hindrance? *Journal for Research in Mathematics Education*, *16*(1), 37–44. https://doi.org/10.2307/748971

Charles A. Dana Center. (2013a). Introduction to The New Mathways Project's Frameworks for Mathematics and Collegiate Learning curriculum (Version 2.0).

Charles A. Dana Center. (2013b). Overview of The New Mathways Project's Foundations of Mathematical Reasoning curriculum.

Charles A. Dana Center. (n.d.). The New Mathways Project. Retrieved October 18, 2015, from http://www.utdanacenter.org/higher-education/new-mathways-project/

Charles A. Dana Center, & Texas Association of Community Colleges. (2014). The New Mathways Project's Foundations of Mathematical Reasoning Curriculum Instructor Materials, Version 2.0.

Clement, J. (2000). Analysis of clinical interviews: Foundations and model viability. In R. Lesh & A. Kelly (Eds.), *Handbook of research methodologies for science and mathematics education* (pp. 341–385). Hillsdale, New Jersey: Lawrence Erlbaum.

Conway, M., & Ross, M. (1984). Getting what you want by revising what you had. *Journal of Personality and Social Psychology*, *47*(4), 738–748. https://doi.org/10.1037/0022-3514.47.4.738

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research*. (V. Knight, S. Connelly, & K. Wiley, Eds.) (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Coulter, S. E. (2012). Using the retrospective pretest to get usable, indirect evidence of student learning. *Assessment & Evaluation in Higher Education*, *37*(3), 321–334. https://doi.org/10.1080/02602938.2010.534761

Cronbach, L. (1961). *Essentials of psychological testing*. New York, NY: Harper and Row.

Denler, H., Wolters, C. A., & Benzon, M. (2014, January 28). Social cognitive theory. Retrieved from http://www.education.com/reference/article/social-cognitive-theory

Detterman, D. K. (1993). The case for the prosecution: Transfer as an epiphenomenon. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 1–24). Norwood, N.J.: Ablex Pub. Corp. Retrieved from http://catalog.hathitrust.org/Record/002627970

Dillman, D., & Christian, L. (2002). *The influence of words, symbols, numbers, and graphics on answers to self-administered questionnaires: Results from 18 experimental comparisons*. Retrieved from http://www.sesrc.wsu.edu/dillman/papers/2002/theinfluencewords.pdf

Dorsey, J., Carvalho, S., & Castillo, A. (2014). The New Mathways Project's curriculum design standards: Selected supporting research. Charles A. Dana Center.

Dorsey, J., Carvalho, S., & Stano, N. (2014). The New Mathways Project's four guiding principles--Selected supporting research. Charles A. Dana Center.

Drennan, J., & Hyde, A. (2008). Controlling response shift bias: The use of the retrospective pre-test design in the evaluation of a master's programme. *Assessment & Evaluation in Higher Education*, *33*(6), 699–709. https://doi.org/10.1080/02602930701773026

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237–251. https://doi.org/10.3102/0013189X15584327

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*(2), 256–273. https://doi.org/10.1037/0033-295X.95.2.256

Dweck, C. S., Chiu, C., & Hong, Y. (1995). Implicit theories and their role in judgments and reactions: A world from two perspectives. *Psychological Inquiry*, *6*(4), 267–285. https://doi.org/10.2307/1448940

Dweck, C. S., Walton, G., & Cohen, G. (2011). *Academic tenacity: Mindsets and skills that promote long-term learning* (White paper prepared for the Gates Foundation). Seattle, WA.

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin*, *21*(3), 215–225. https://doi.org/10.1177/0146167295213003

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, *113*(28), 7900–7905. https://doi.org/10.1073/pnas.1602413113

Elkind, D. (1964). Piaget's semi-clinical interview and the study of spontaneous religion. *Journal for the Scientific Study of Religion*, *4*(1), 40–47.

Endler, N. S., & Kocovski, N. L. (2001). State and trait anxiety revisited. *Journal of Anxiety Disorders*, *15*(3), 231–245. https://doi.org/10.1016/S0887-6185(01)00060-3

Farrington, C., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T., Johnson, D., & Beechum, N. (2012). *Teaching adolescents to become learners: The role of noncognitive factors in shaping school performance—A critical literature review*. Chicago, IL: University of Chicago Consortium on Chicago School Research.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Faust, M. W., Ashcraft, M. H., & Fleck, D. E. (1996). Mathematics anxiety effects in simple and complex addition. *Mathematical Cognition*, *2*, 25–62.

Garcia, E. (2014). *The need to address noncognitive skills in the education policy agenda* (Policy brief 386). Washington, D.C.: Economic Policy Institute. Retrieved from http://s3.epi.org/files/2014/the-need-to-address-noncognitive-skills-12-02-2014.pdf

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1–38.

Ginsburg, H. P. (1981). The clinical interview in psychological research on mathematical thinking: Aims, rationales, techniques. *For the Learning of Mathematics*, *1*(3), 4–11.

Ginsburg, H. P. (1997). *Entering the child's mind: The clinical interview in psychological research and practice*. New York, NY: Cambridge University Press.

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2004). Collective efficacy beliefs: Theoretical developments, empirical evidence, and future directions. *Educational Researcher*, *33*(3), 3–13. https://doi.org/10.3102/0013189X033003003

Golembiewski, R. T., Billingsley, K., & Yeager, S. (1976). Measuring change and persistence in human affairs: Types of change generated by OD designs. *The Journal of Applied Behavioral Science*, *12*(2), 133–157. https://doi.org/10.1177/002188637601200201

Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology*, *102*(4), 700–717. https://doi.org/10.1037/a0026659

Gunderson, E., Ramirez, G., Levine, S., & Beilock, S. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles*, *66*(3–4), 153–166. https://doi.org/10.1007/s11199-011-9996-2

Hannigan, M. K. A. (1998). *Exploration of an instructional strategy to promote explicit understanding of place value concepts in prospective elementary teachers* (Ph.D.). The University of Texas at Austin, Ann Arbor. Retrieved from ProQuest Dissertations & Theses Full Text. (304457642)

Hatano, G. (1988). Social and motivational bases for mathematical understanding. *New Directions for Child and Adolescent Development*, *41*, 55–70.

Hatano, G., & Inagaki, K. (1986). Two courses in expertise. In H. Stevenson, J. Azuma, & K. Hakuta (Eds.), *Child development and education in Japan*. New York, NY: W. H. Freeman & Co.

Hatano, G., & Oura, Y. (2003). Commentary: Reconceptualizing school learning using insight from expertise research. *Educational Researcher*, *32*(8), 26–29. https://doi.org/10.2307/3700083

Hausmann, L., Schofield, J., & Woods, R. (2007). Sense of belonging as a predictor of intentions to persist among African American and white first-year college students. *Research in Higher Education*, *48*(7), 80–839.

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, *21*(1), 33–46. https://doi.org/10.2307/749455

Hill, L. G., & Betz, D. L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation*, *26*(4), 501–517. https://doi.org/10.1177/1098214005281356

Hoogstraten, J. (1982). The retrospective pretest in an educational training context. *The Journal of Experimental Education*, *50*(4), 200–204. https://doi.org/10.2307/20151460

Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review*, *4*(1), 93–106. https://doi.org/10.1177/0193841X8000400105

Howard, G. S., Dailey, P. R., & Gulanick, N. A. (1979). The feasibility of informed pretests in attenuating response-shift bias. *Applied Psychological Measurement*, *3*(4), 481–494. https://doi.org/10.1177/014662167900300406

Howard, G. S., Millham, J., Slaten, S., & O'Donnell, L. (1981). Influence of subject response style effects on retrospective measures. *Applied Psychological Measurement*, *5*(1), 89–100. https://doi.org/10.1177/014662168100500113

Howard, G. S., Ralph, K. M., Gulanick, N. A., Maxwell, S. E., Nance, D. W., & Gerber, S. K. (1979). Internal invalidity in pretest-posttest self-report evaluations and re-evaluation of retrospective pretests. *Applied Psychological Measurement*, *3*(1), 1–23. https://doi.org/10.1177/014662167900300101

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly*, *14*(3), 299–324. https://doi.org/10.1111/j.1471-6402.1990.tb00022.x

Jameson, M. M., & Fusco, B. R. (2014). Math anxiety, math self-concept, and math self-efficacy in adult learners compared to traditional undergraduate students. *Adult Education Quarterly*, *64*(4), 306–322. https://doi.org/10.1177/0741713614541461

Jenkins, D., & Sung-Woo, C. (2014). *Get with the program...and finish it: Building guided pathways to accelerate student completion* (Working Paper No. 66). Community College Research Center, Teachers College, Columbia University. Retrieved from http://ccrc.tc.columbia.edu/publications/get-with-the-program-finish-it.html

Johnstone, R. (2015). *Guided pathways demystified: Exploring ten commonly asked questions about implementing pathways*. National Center for Inquiry and Improvement. Retrieved from http://www.inquiry2improvement.com

Kamins, M. L., & Dweck, C. S. (1999). Person versus process praise and criticism: Implications for contingent self-worth and coping. *Developmental Psychology*, *35*(3), 835.

Kivel, L. (2014, September 11). Creating opportunities for students to become flexible experts. Retrieved from http://www.carnegiefoundation.org/blog/creating-opportunities-students-become-flexible-experts

Klatt, J., & Taylor-Powell, E. (2005, October). *Synthesis of literature relative to the retrospective pretest design*. Paper presented at the Joint CES/AEA Conference, Toronto, Canada.

Kohn, A. (2015, August 16). The education fad that's hurting our kids: What you need to know about "Growth Mindset" theory — and the harmful lessons it imparts. *Salon*. Retrieved from http://www.salon.com/2015/08/16/the_education_fad_thats_hurting_our_kids_what_you_need_to_know_about_growth_mindset_theory_and_the_harmful_lessons_it_imparts

Kosovich, J. J., Hulleman, C. S., Barron, K. E., & Getty, S. (2014). A practical measure of student motivation: Establishing validity evidence for the expectancy-value-cost scale in middle school. *The Journal of Early Adolescence*. https://doi.org/10.1177/0272431614556890

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, *50*, 537–567.

Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In P. Marsden & J. Wright (Eds.), *Handbook of Survey Research* (2nd ed., pp. 263–313).

Emerald. Retrieved from http://studysites.sagepub.com/kumar4e/study/Chapter%209/Questionnaires.pdf

Lam, T. C., & Bengo, P. (2003). A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation*, *24*(1), 65–80. https://doi.org/10.1177/109821400302400106

Lee, V., Russ, R., & Sherin, B. (2008). A functional taxonomy of discourse moves for conversation management during cognitive clinical interviews about scientific phenomena. In V. Sloutsky, B. Love, & K. McRae (Eds.), *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (pp. 1723–1728). Austin, Texas.

Leech, B. (2002). Asking questions: Techniques for semistructured interviews. *Political Science and Politics*, *35*(4), 665–668. https://doi.org/10.1017/S1049096502001129

Lewis, A. (1970). The ambiguous word“ anxiety”. *International Journal of Psychiatry*, *9*, 62–79.

Liebert, R., & Morris, L. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports*, *20*, 975–978. https://doi.org/10.2466/pr0.1967.20.3.975

Linnenbrink, E. A., & Pintrich, P. R. (2002). Motivation as an enabler for academic success. *School Psychology Review*, *31*(3), 313–327.

Lobato, J. (2006). Alternative perspectives on the transfer of learning: History, issues, and challenges for future research. *The Journal of the Learning Sciences*, *15*(4), 431–449. https://doi.org/10.1207/s15327809jls1504_1

Lyons, I. M., & Beilock, S. L. (2012). When math hurts: Math anxiety predicts pain network activation in anticipation of doing math. *PLoS ONE*, *7*(10), e48076. https://doi.org/10.1371/journal.pone.0048076

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education*, *30*(5), 520–540. https://doi.org/10.2307/749772

Maloney, E. A., & Beilock, S. L. (2012). Math anxiety: Who has it, why it develops, and how to guard against it. *Trends in Cognitive Sciences*, *16*(8), 404–406. https://doi.org/10.1016/j.tics.2012.06.008

Martin, T., Baker Peacock, S., Ko, P., & Rudolph, J. (2015). Changes in teachers' adaptive expertise in an engineering professional development course. *Journal of Pre-College Engineering Education Research*, *5*(2), 1–14. https://doi.org/10.7771/2157-9288.1050

Mathematical Association of America's Committee on the Undergraduate Program in Mathematics. (2015). *2015 Committee on the Undergraduate Program in Mathematics' curriculum guide to majors in mathematical sciences*. Mathematical Association of America.

Meuschke, D. M. (2005). *The relationship between goal-orientation, help-seeking, math self-efficacy, and mathematics achievement in a community college* (Ed.D.). University of Southern California, Ann Arbor. Retrieved from ProQuest Dissertations & Theses Full Text. (305426246)

Mezoff, B. (1981). How to get accurate self-reports of training outcomes. *Training & Development Journal*, *35*(9), 56.

Midgley, C., Maehr, M. L., Hruda, L., Anderman, E., Anderman, L. H., Freeman, K., … Urdan, T. (2000). Manual for the Patterns of Adaptive Learning Scales (PALS). University of Michigan.

Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology*, *75*(1), 33–52. https://doi.org/10.1037/0022-3514.75.1.33

National Govenors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*. Washington, D.C.

National Research Council. (2000). Learning and transfer. In J. D. Bransford, A. L. Brown, & R. R. Cocking (Eds.), *How people learn: Brain, mind, experience, and school (Expanded edition)* (pp. 31–78). Washington, D.C.: National Academy Press.

Niiya, Y., Brook, A. T., & Crocker, J. (2010). Contingent self-worth and self-handicapping: Do incremental theorists protect self-esteem? *Self and Identity*, *9*(3), 276–297. https://doi.org/10.1080/15298860903054233

Nimon, K. (2007). *Comparing outcome measures derived from four research designs incorporating the retrospective pretest* (Ph.D.). University of North Texas, Denton, Texas. Retrieved from ProQuest Dissertations & Theses Full Text.

Nimon, K. (2014). Explaining differences between retrospective and traditional pretest self-assessments: competing theories and empirical evidence. *International Journal of Research & Method in Education*, *37*(3), 256–269. https://doi.org/10.1080/1743727X.2013.820644

Nimon, K., Zigarmi, D., & Allen, J. (2011a). Measures of program effectiveness based on retrospective pretest data: Are all created equal? *American Journal of Evaluation*, *32*(1), 8–28. https://doi.org/10.1177/1098214010378354

Nimon, K., Zigarmi, D., & Allen, J. (2011b). Measures of program effectiveness based on retrospective pretest data: Are all created equal? *American Journal of Evaluation*, *32*(1), 8–28. https://doi.org/10.1177/1098214010378354

Norman, G. (2003). Hi! How are you? Response shift, implicit theories and differing epistemologies. *Quality of Life Research*, *12*(3), 239–249. https://doi.org/10.1023/A:1023211129926

Packer, M. (2001). The problem of transfer, and the sociocultural critique of schooling. *Journal of the Learning Sciences*, *10*(4), 493–514. https://doi.org/10.1207/S15327809JLS1004new_4

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*(4), 543–578. https://doi.org/10.3102/00346543066004543

Pajares, F. (1997). Current directions in self-efficacy research. *Advances in Motivation and Achievement*, *10*(149), 1–49.

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology*, *86*(2), 193–203. https://doi.org/10.1037/0022-0663.86.2.193

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40. https://doi.org/10.1037/0022-0663.82.1.33

Pomerantz, E. M., & Kempner, S. G. (2013). Mothers' daily person and process praise: Implications for children's theory of intelligence and motivation. *Developmental Psychology*, *49*(11), 2040–2046. https://doi.org/10.1037/a0031840

Ramirez, G., & Beilock, S. L. (2011). Writing about testing worries boosts exam performance in the classroom. *Science*, *331*, 211–213. https://doi.org/10.1126/science.1199427

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *Journal of Cognition and Development*, *14*(2), 187–202. https://doi.org/10.1080/15248372.2012.664593

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review*, *96*(2), 341–357. https://doi.org/10.1037/0033-295X.96.2.341

Ross, M. E., Shannon, D. M., Salisbury-Glennon, J. D., & Guarino, A. (2002). The Patterns of Adaptive Learning Survey: A comparison across grade levels. *Educational and Psychological Measurement*, *62*(3), 483–497. https://doi.org/10.1177/00164402062003006

Rudestam, K., & Newton, R. (1992). *Surviving your dissertation: A comprehensive guide to content and process* (1st ed.). Sage Publications, Inc.

Rusch, T. L. (1997). *Mathematics content coursework for prospective elementary teachers: Examining the influence of instructional strategy on the development of essential place value knowledge* (Ph.D.). The University of Texas at Austin, Ann Arbor. Retrieved from ProQuest Dissertations & Theses Full Text. (304372092)

Rutschow, E. Z., & Diamond, J. (2015). *Laying the foundations: Early findings from The New Mathways Project*. MDRC. Retrieved from http://www.mdrc.org/publication/laying-foundations

Saris, W. E., Revilla, M., Krosnick, J. A., & Shaeffer, E. M. (2010). Comparing questions with agree/disagree response options to questions with item-specific response options. In *Survey Research Methods* (Vol. 4, pp. 61–79). Retrieved from http: // www.surveymethods.org

Schunk, D. H. (1981). Modeling and attributional effects on children's achievement: A self-efficacy analysis. *Journal of Educational Psychology*, *73*(1), 93.

Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist*, *26*(3–4), 207–231.

Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. *Transfer of Learning from a Modern Multidisciplinary Perspective*, 1–51.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, *22*(2), 129–184. https://doi.org/10.1207/s1532690xci2202_1

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*(2), 93–105.

Shell, D., Murphy, C., & Bruning, R. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology*, *81*, 91–100.

Spradley, J. (1979). *The ethnographic interview*. New York, NY: Harcourt, Brace, Jovanovich.

Sprangers, M. A. (1996). Response-shift bias: A challenge to the assessment of patients' quality of life in cancer clinical trials. *Goals of Palliative Cancer Therapy II*, *22*, *Supplement A*, 55–62. https://doi.org/10.1016/S0305-7372(96)90064-X

Sprangers, M. A., & Hoogstraten, J. (1989). Pretesting effects in retrospective pretest-posttest designs. *Journal of Applied Psychology*, *74*(2), 265–272. https://doi.org/10.1037/0021-9010.74.2.265

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613–629. https://doi.org/10.1037/0003-066X.52.6.613

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797–811. https://doi.org/10.1037/0022-3514.69.5.797

Stigler, J., Givvin, K., & Thompson, B. (2010). *What community college developmental mathematics students understand about mathematics* (Exploration Paper). Los Angeles, California: University of California. Retrieved from www.carnegiefoundation.org/elibrary/problem-solution-exploration-papers

Taminiau-Bloem, E. F., Schwartz, C. E., van Zuuren, F. J., Koeneman, M. A., Visser, M. R. M., Tishelman, C., … Sprangers, M. A. G. (2016). Using a retrospective pretest instead of a conventional pretest is replacing biases: A qualitative study of cognitive processes underlying responses to thentest items. *Quality of Life Research*, *25*, 1327–1337. https://doi.org/10.1007/s11136-015-1175-4

The Texas Association of Community Colleges. (2012). *New initiative aims to advance developmental mathematics students in Texas community colleges* (Press release). Retrieved from http://www.tacc.org

Tourangeau, R. (2004). Survey research and societal change. *Annual Review of Psychology*, *55*(1), 775–801. https://doi.org/10.1146/annurev.psych.55.090902.142040

Treisman, U. (2015, November 13). The New Mathways Project and National Coalition Release Core Principles for Student Success. Retrieved from http://us2.campaign-archive2.com/?u=f75754127932b3bd8ffbea25c&id=e96ef7028e&e=[UNIQID]

Uebersax, J. (2006). Likert scales: Dispelling the confusion. Retrieved June 28, 2016, from http://www.john-uebersax.com/stat/likert.htm

Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology*, *34*(1), 89–101. https://doi.org/10.1016/j.cedpsych.2008.09.002

van Aalderen- Smeets, S. I., Walma van der Molen, J. H., & Asma, L. J. (2012). Primary teachers' attitudes toward science: A new theoretical framework. *Science Education*, *96*(1), 158–182. https://doi.org/10.1002/sce.20467

Vandal, B. (2015, February). *Completing gateway courses and entering programs of study*. Webinar presented at the Creating a college-ready pipeline to post-secondary programs, Washington, D.C. Retrieved from https://vimeo.com/thenrocproject/review/120187245/e40986c9a0

Walton, G., & Carr, P. B. (2011). Social belonging and the motivation and intellectual achievement of negatively stereotyped students. In M. Inzlicht & T. Schmader (Eds.), *Stereotype threat: Theory, process, and application* (pp. 89–106). Oxford University Press.

Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science*, *331*(6023), 1447–1451. https://doi.org/10.1126/science.1198364

Walton, G. M., Cohen, G. L., Cwir, D., & Spencer, S. J. (2012). Mere belonging: The power of social connections. *Journal of Personality and Social Psychology*, *102*(3), 513–532. https://doi.org/10.1037/a0025731

West, M., Kraft, M., Finn, A., Martin, R., Duckworth, A. L., Gabrieli, C., & Gabrieli, J. (2014). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. Presented at the CESifo Area Conference on Economics of Education, Munich, Germany.

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Woltman, H., Feldstain, A., MacKay, C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 52–69.

Yeager, D. S., Bryk, A., Muhich, J., Hausman, H., & Morales, L. (2013). *Practical measurement*. Carnegie Foundation for the Advancement of Teaching. Retrieved from http://cdn.carnegiefoundation.org/wp-content/uploads/2014/09/Practical_Measurement_Yeager-Bryk1.pdf

Yeager, D. S., Paunesku, D., Walton, G. M., & Dweck, C. S. (2013). *How can we instill productive mindsets at scale? A review of the evidence and an initial R&D agenda* (White paper prepared for the White House meeting on "Excellence in education: The importance of academic mindsets").

Yeager, D. S., Walton, G., & Cohen, G. L. (2013). Addressing achievement gaps with psychological interventions. *Phi Delta Kappan*, *94*(5), 62–65.

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research*, *81*(2), 267–301. https://doi.org/10.3102/0034654311405999

Young, C. B., Wu, S. S., & Menon, V. (2012). The neurodevelopmental basis of math anxiety. *Psychological Science*, *23*(5), 492–501. https://doi.org/10.1177/0956797611429134

Zientek, L. R., Yetkiner, Z. E., & Thompson, B. (2010). Characterizing the mathematics anxiety literature using confidence intervals as a literature review mechanism. *The Journal of Educational Research*, *103*(6), 424–438. https://doi.org/10.1080/00220670903383093

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, *25*(1), 82–91. https://doi.org/10.1006/ceps.1999.1016

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166–183. https://doi.org/10.3102/0002831207312909